

PHILOSOPHY OF SCIENCE

THE CENTRAL ISSUES

MARTIN CURD & J. A. COVER

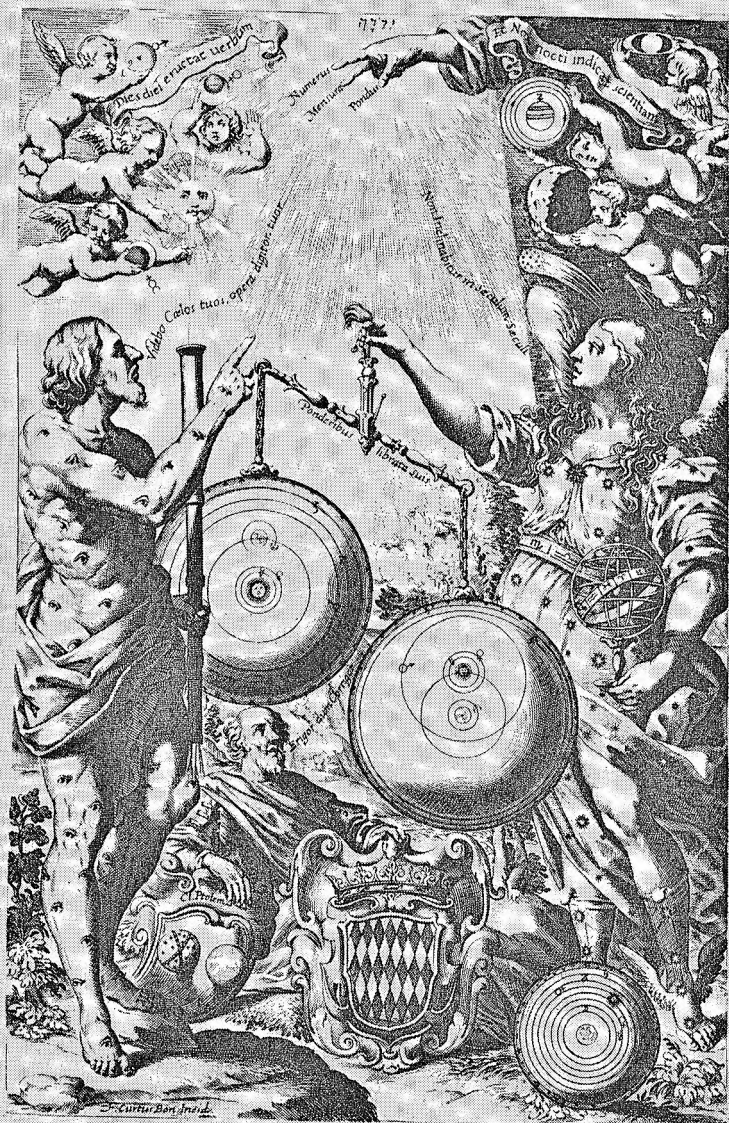
| Philosophy of Science

ABOUT THE FRONTISPIECE

The frontispiece is from *Almagestum Novum* (Bologna, 1651), by the Jesuit astronomer Giambattista Riccioli (1598–1671). In the decades following the condemnation of Galileo, Riccioli was an ardent critic of the Copernican theory. He conceded that Galileo's discovery of the phases of Venus had refuted the Ptolemaic system but insisted that Tycho Brahe's system, in which the earth does not move, captured all the observational and mathematical advantages of the Copernican theory with none of its physical and theological disadvantages. Riccioli's book (whose title is a deliberate reference to the "old" *Almagest* of Ptolemy, now discredited) gives an exhaustive survey of arguments for and against the Copernican theory, and concludes that Tycho Brahe's system (modified slightly by Riccioli) is more plausible.

Thus, Riccioli's frontispiece shows his own version of the Tychonic system weighing more heavily in the scales of evidence than its Copernican rival. In Riccioli's variant, Mercury, Venus, and Mars are satellites of the sun, but, unlike Brahe's original scheme, Jupiter and Saturn are centered on the earth. The figure holding the scales and the armillary sphere combines features of Urania (the muse of astronomy) and Astraea (the goddess of justice). On the left is hundred-eyed Argus, observing the sun through a telescope held to an eye on his knee. His words allude to Psalm 8, verse 3: "When I consider thy heavens, the work of thy fingers. . . ." At the bottom lies Ptolemy with his discarded system. Ptolemy rests his hand on the coat of arms of the prince of Monaco (to whom the *Almagestum Novum* was dedicated) magnanimously acknowledging the correction of his errors. At the top are depicted recent astronomical discoveries of the seventeenth century: Mercury and Venus displaying crescent phases; Saturn with two "handles"—this was prior to Huyghens's ring hypothesis; Jupiter with four moons and two bands parallel to its equator (a feature first noted by Riccioli); a heavily cratered moon; and a comet soaring through the heavens like a spotted cannonball. In the center at the top is the Hebrew word *Yah-Veh* and a reference to the Wisdom of Solomon 11, verse 20: "But thou hast ordered all things by measure and number and weight." On the left and right are quotations from Psalm 19, verse 2: "Day unto day uttereth speech, and night unto night sheweth knowledge."

Although Riccioli's book had no effect on the debate over the Copernican theory—by the middle of the seventeenth century, almost all scientists and astronomers were Copernicans—it illustrates one of the most important contests between rival theories in the history of science.



Sunt ex libris Petri à Sacerdotis
L. Bat.

Philosophy of Science

THE CENTRAL ISSUES

Martin Curd | J. A. Cover



W. W. Norton & Company

NEW YORK | LONDON

Copyright © 1998 by W. W. Norton & Company, Inc.

All rights reserved

Printed in the United States of America

First Edition

The text of this book is composed in **Electra LH**
with the display set in **Electra LH**.

Composition by PennSet, Inc.

Manufacturing by Courier, Westford

Book design by Martin Lubin Graphic Design

Permissions acknowledgments appear on pages 1331–34, which constitute a
continuation of the copyright page

Library of Congress Cataloging-in-Publication Data

Philosophy of science : the central issues / Martin Curd, J. A. Cover,
[editors].

p. cm.

Includes bibliographical references and index.

ISBN 0-393-97175-9 (pbk.)

I. Science — Philosophy. I. Curd, Martin. II. Cover, J. A. (Jan
A.), 1958– .

Q175 P5129 1998

501—dc21

97-39387

W. W. Norton & Company, Inc., 500 Fifth Avenue, New York, N.Y. 10110
<http://www.wwnorton.com>

W. W. Norton & Company Ltd.

Castle House, 75/76 Wells Street, London W1T 3QT

1 2 3 4 5 6 7 8 9 0

To the memory of
Carl G. Hempel (1905–1997)
Thomas S. Kuhn (1922–1996)

CONTENTS

Preface xv

General Introduction xvii

1 | Science and Pseudoscience

INTRODUCTION 1

Karl Popper

Science: Conjectures and Refutations 3

Thomas S. Kuhn

Logic of Discovery or Psychology of Research? 11

Imre Lakatos

Science and Pseudoscience 20

Paul R. Thagard

Why Astrology Is a Pseudoscience 27

Michael Ruse

Creation-Science Is Not Science 38

Larry Laudan

Commentary: Science at the Bar—Causes for Concern 48

Michael Ruse

Response to the Commentary: *Pro Judice* 54

COMMENTARY 62

2 | Rationality, Objectivity, and Values in Science

INTRODUCTION 83

Thomas S. Kuhn

The Nature and Necessity of Scientific Revolutions 86

<i>Thomas S. Kuhn</i>	
Objectivity, Value Judgment, and Theory Choice	102
<i>Ernan McMullin</i>	
Rationality and Paradigm Change in Science	119
<i>Larry Laudan</i>	
Dissecting the Holist Picture of Scientific Change	139
<i>Helen E. Longino</i>	
Values and Objectivity	170
<i>Kathleen Okruhlik</i>	
Gender and the Biological Sciences	192
COMMENTARY	210

3 | The Duhem-Quine Thesis and Underdetermination

INTRODUCTION	255
<i>Pierre Duhem</i>	
Physical Theory and Experiment	257
<i>W. V. Quine</i>	
Two Dogmas of Empiricism	280
<i>Donald Gillies</i>	
The Duhem Thesis and the Quine Thesis	302
<i>Larry Laudan</i>	
Demystifying Underdetermination	320
COMMENTARY	354

4 | Induction, Prediction, and Evidence

INTRODUCTION	409
<i>Peter Lipton</i>	
Induction	412
<i>Karl Popper</i>	
The Problem of Induction	426

Wesley C. Salmon

Rational Prediction 433

Carl G. Hempel

Criteria of Confirmation and Acceptability 445

Laura J. Snyder

Is Evidence Historical? 460

Peter Achinstein

Explanation v. Prediction: Which Carries More Weight? 481

COMMENTARY 494

5 | Confirmation and Relevance: Bayesian Approaches

INTRODUCTION 549

Wesley C. Salmon

**Rationality and Objectivity in Science or Tom Kuhn Meets
Tom Bayes 551**

Clark Glymour

Why I Am Not a Bayesian 584

Paul Horwich

Wittgensteinian Bayesianism 607

COMMENTARY 626

6 | Models of Explanation

INTRODUCTION 675

Rudolf Carnap

The Value of Laws: Explanation and Prediction 678

Carl G. Hempel

Two Basic Types of Scientific Explanation 685

Carl G. Hempel

The Thesis of Structural Identity 695

Carl G. Hempel

Inductive-Statistical Explanation 706

David-Hillel Ruben

Arguments, Laws, and Explanation 720

Peter Railton

A Deductive-Nomological Model of Probabilistic Explanation 746

COMMENTARY 766

7 | Laws of Nature

INTRODUCTION 805

A. J. Ayer

What Is a Law of Nature? 808

Fred I. Dretske

Laws of Nature 826

D. H. Mellor

Necessities and Universals in Natural Laws 846

Nancy Cartwright

Do the Laws of Physics State the Facts? 865

COMMENTARY 878

8 | Intertheoretic Reduction

INTRODUCTION 903

Ernest Nagel

Issues in the Logic of Reductive Explanations 905

Paul K. Feyerabend

**How to Be a Good Empiricist—A Plea for Tolerance
in Matters Epistemological 922**

Thomas Nickles

Two Concepts of Intertheoretic Reduction 950

Philip Kitcher

1953 and All That: A Tale of Two Sciences 971

COMMENTARY 1004

9 | Empiricism and Scientific Realism

INTRODUCTION 1049

Grover Maxwell

The Ontological Status of Theoretical Entities 1052

Bas C. van Fraassen

Arguments Concerning Scientific Realism 1064

Alan Musgrave

Realism versus Constructive Empiricism 1088

Larry Laudan

A Confutation of Convergent Realism 1114

James Robert Brown

Explaining the Success of Science 1136

Ian Hacking

Experimentation and Scientific Realism 1153

David B. Resnik

Hacking's Experimental Realism 1169

Arthur Fine

The Natural Ontological Attitude 1186

Alan Musgrave

NOA's Ark—Fine for Realism 1209

COMMENTARY 1226

Glossary 1291

Bibliography 1311

Permissions Acknowledgments 1331

Name Index 1335

Subject Index 1349

PREFACE

Many people have helped us during the writing and preparation of this book. We are especially indebted to the philosophers who commented at length on earlier versions of the manuscript. As these reviewers can now verify, their criticisms and suggestions have played a major role in shaping the book's final form. They are Marthe Chandler, Len Clark, Keith J. Cooper, Richard Creath, Evan Fales, Stuart S. Glennan, Paul Humphreys, David Magnus, Deborah G. Mayo, and Bonnie Paller. We are also grateful to Rod Bertolet, Clark Glymour, Bill Gustason, Dick Jeffrey, Dana Mason, Alan Musgrave, Tom Nickles, Kathleen Okruhlik, and Ted Ulrich for helping us on topics ranging from Ramsey's dot notation and three-valued logic to T. S. Eliot's poetry, pig farming, and the aetiology of AIDS. A special thanks is due to our secretary at Purdue, Pamela Connelly, and to our editor at Norton, Allen Clawson, for seeing this project through with skill, enthusiasm, and good humor. Without Allen, the book would not have been begun and, almost certainly, would not have been completed. The staff at Norton have been exemplary. We would particularly like to acknowledge Jane Carter (for her meticulous manuscript editing), Claire Acher (for obtaining the permissions and preparing the manuscript for production), and Roy Tedoff (who guided the entire production process). Finally, there are our wives, Patricia and Karen, to whom any formal expression of thanks seems inadequate.

GENERAL INTRODUCTION

The philosophy of science is at least as old as Aristotle, but it has risen to special prominence in the twentieth century. As scientists have made tremendous advances in fields as diverse as genetics, geology, and quantum mechanics, increasing numbers of philosophers have made science their focus of study. In its broadest terms, the philosophy of science is the investigation of philosophical questions that arise from reflecting on science. What makes these questions philosophical is their generality, their fundamental character, and their resistance to solution by empirical disciplines such as history, sociology, and psychology.

The difference between the philosophy of science and other disciplines that study science can be brought out by contrasting different sorts of question. For example, "When was the planet Neptune discovered?" is primarily a question for historians, not for philosophers.¹ Similarly, "Why did Soviet biologists under Stalin reject Mendelian genetics?" or "Why did James Watson underrate the contributions of Rosalind Franklin to the work that led to the discovery of the double helix structure of DNA?" fall within the domains of sociology, political science, and psychology. Contrast these questions with the following: "When is a theory confirmed by its predictions?" "Should we be realists about all aspects of well-established theories?" "What is a law of nature?" These questions are philosophical. They cannot be answered simply by finding out what has happened in the past or what people now believe.

For similar reasons, philosophical questions about science cannot be answered by the sciences themselves (although being able to answer these questions often depends on having a good understanding of scientific theories). A geneticist at the National Cancer Institute, for example, might ask whether certain people are born with a natural immunity to AIDS and set out to answer this question through empirical research. But if our geneticist asked "What is a law of nature?" or "What is science?" or "When is a theory confirmed?" she would not discover the answer by doing more science.

The central questions in the philosophy of science do not belong to science as such; they are *about* science, but not *part of it*. Of course, scientists can be (and sometimes have been) philosophers of science. The point is that when people are doing philosophy of science, they are not (usually) doing science per se, and most philosophers of science (at least in the twentieth century) have not been practicing scientists. Thus, the philosophy of science is not a branch of science but belongs to philosophy, and it intersects with other areas of philosophy, such as epistemology, metaphysics, and the philosophy of language.

The aim of *Philosophy of Science: The Central Issues* is to introduce

the reader to the main currents in twentieth-century philosophy of science. It is primarily intended for use in introductory courses at both the undergraduate and graduate levels. In order to keep the book within manageable bounds, some difficult decisions had to be made about what to include and what to exclude. In making these decisions we were guided by our own experience in teaching at Purdue and the recommendations of our reviewers who contributed significantly to the book's development.

The first and in some ways the easiest decision was to exclude the social sciences and concentrate exclusively on the natural sciences. In this we followed the lead of other texts. The philosophical questions raised by disciplines such as history, psychology, sociology, and anthropology are fascinating and important. But they are so different from the questions one encounters in physics, biology, and chemistry that they would require another volume, comparable in length to this one, in order to address them adequately.

A second decision, which we made at the outset, was to avoid foundational questions about the concepts, structure, and content of particular theories and to focus instead on general issues that arise across scientific disciplines. Thus, this volume is organized around wide-ranging philosophical topics and problems, not individual theories or sciences. Details of particular sciences are introduced rarely and only when necessary for evaluating a philosophical position or argument (as, for example, in the chapter on reduction). In this way we hope to avoid the trap of turning a philosophy of science course into a minicourse in science and to keep the focus on the *philosophy* in the philosophy of science. It also has the advantage of making courses based on this book accessible to students (even those at the graduate level) whose background in the sciences may be slight or nonexistent. For the same reason of accessibility, we have confined our selections to readings that use no more than a bare minimum of logical or mathematical notation. The one place where a certain amount of formal notation is unavoidable is in chapter 5, on Bayesian approaches to confirmation theory. But even there, we have edited the readings (sometimes by adding an editorial footnote, sometimes by changing the notation) in order to make them easier to understand, and we have provided an introduction to Bayes's theorem and the probability calculus in the accompanying commentary.

Our approach, then, is focused on philosophical topics and problems, not on particular sciences and theories. A consequence of this topics-and-problems approach is that the chapters of *Philosophy of Science: The Central Issues* pay little attention to tracing the historical development of the philosophy of science in twentieth century. Although we devote some time to filling in some of the essential historical background in the commentaries, this volume is not historical in the way that it treats ideas, arguments, or philosophers. What connects the readings (and the discussions of them in the commentaries) is their focus on common themes, argu-

ments, and criticisms, regardless of whether the authors share the same nationality, are writing in the same decade, or belong to the same school. Our approach is not antihistorical, but it is largely ahistorical.

Because of the many sharp disagreements within the philosophy of science and the unresolved character of nearly all the fundamental questions that philosophers ask about science, an anthology seemed to us to be the only sane choice for a book intended for use in the classroom. But the anthology format brings with it a problem that just about every teacher of philosophy of science has had to confront. Hardly any of the readings, whether old classics or brand-new articles, were written with students in mind. Rather, they were published in books and professional journals, addressed primarily to fellow professionals. Thus, they often presuppose an awareness of issues, positions, and arguments, both in the philosophy of science and in philosophy more generally, that most students lack. Consequently, even the brightest students can find it hard to understand the material they are being asked to read, discuss, and evaluate. The most common complaint voiced by the teachers we spoke with in the several years that went into planning and writing this book, is that many of the readings in the existing anthologies are too sophisticated—they make too many references to the history of science and allude too frequently to philosophical ideas and arguments for the beginning student to get much out of them. What was needed, and what we have tried to provide here, is a serious, comprehensive guide that will really help students in their first encounter with the readings. Thus, in addition to short introductions to each chapter, we have written extended and often detailed commentaries on the readings. Getting the tone and level of detail right in these commentaries has been the hardest and most rewarding part of the book's development. Much of the fine tuning and, in some cases, the inclusion and deletion of entire sections, was guided by our reviewers. We have strived to make each commentary and the sections within them self-contained so that each can be used independently of the rest. And in order to maximize the pedagogical usefulness of *Philosophy of Science: The Central Issues*, each reading is linked explicitly with one or more of the sections into which the commentaries are divided. In this way, where one should look in the commentaries for discussion, explanation, background, and analysis of any of the forty-nine separate readings in the book should be clear.

At the end of this volume there is a glossary, a bibliography, and indexes of names and subjects. The glossary is comprehensive: it covers most of the terms that may be new to the reader or that are being used in an unfamiliar way. The bibliography is divided into nine sections, one for each chapter. Inevitably, this involves some repetition of titles of books and articles, but our aim was to provide the reader with suggestions for further reading, at an appropriate level, about the issues discussed in each chapter's commentary. Consequently, not everything cited in the com-

mentaries appears in the bibliography, some items appear in the bibliography more than once, and there are some things in the bibliography that are not mentioned in the commentaries

The difference between an anthology and a heap of articles lies in their organization. But any system of division will be, to some extent, artificial and misleading: artificial because of the interrelated character of the issues in the philosophy of science and misleading because it might suggest that the readings in one chapter are not connected with those in another. Thus, as with any collection of this kind, the reader or teacher needs to bear in mind that not everything pertinent to, say, the topic of laws will be found in the chapter devoted to laws and that relevant readings and commentaries might also appear in the chapters on explanation and confirmation (as indeed, in this case, they do). Moreover, this is a collection of readings on related topics, not an extended narrative with a beginning, a middle, and an end. Users of the book should not feel constrained by the order of the chapters or even, in most cases, by the order of the readings within those chapters, when deciding what to read first, what to read second, and so on. Obviously, we have arranged the material in an order that makes sense to us, trying wherever possible to juxtapose readings that speak to the same or closely related issues, but many different arrangements are possible and may be preferable, depending on one's interests and teaching goals.

■ | Notes

1. This is not to deny that the question might raise philosophical issues concerning the concept of discovery. For example, suppose that an astronomer takes a photograph of the night sky through a telescope that is powerful enough to render Neptune visible. Up to that time, no one has seen Neptune. When the plate is developed, it contains an image of Neptune. Although the astronomer sees the image and records the position of the body which made it, he believes that it is "just another star," not a new planet. Has the astronomer discovered Neptune?

| Philosophy of Science



I | Science and Pseudoscience

INTRODUCTION

Parapsychology is defined by its practitioners as the study of extrasensory perception (ESP) and paranormal powers such as telekinesis. ESP includes such alleged psychic phenomena as telepathy, clairvoyance, and precognition. Shunned for decades by the scientific establishment, parapsychologists received official recognition in 1969 when the American Association for the Advancement of Science (the AAAS) admitted the Parapsychological Association as an affiliate member. Many scientists are unhappy with this decision, since they regard parapsychology as a pseudoscience. In 1979, the renowned physicist John A. Wheeler wrote a blistering letter to the president of the AAAS urging that the parapsychologists be expelled from the association. Wheeler wrote, "We have enough charlatanism in this country today without needing a scientific organization to prostitute itself to it. The AAAS has to make up its mind whether it is seeking popularity or whether it is strictly a scientific organization."¹

The debate about the nature of science—about its scope, methods, and aims—is as old as science itself. But this debate becomes especially heated when one group of practitioners accuses another group of practicing pseudoscience. In the twentieth century many individuals, groups, and theories have been accused of being pseudoscientific, including Freud and psychoanalysis, astrology, believers in the paranormal, Immanuel Velikovsky and Erich von Daniken (whose best-selling books *Worlds in Collision* and *Chariots of the Gods* excited the wrath of Carl Sagan and the scientific establishment), and, most recently, the self-styled advocates of creation-science. The proponents of astrology, the paranormal, psychoanalysis, and creation-science engage in research, write books, and publish articles, but their work is typically found in popular magazines and bookstores rather than refereed journals and science libraries. They are seldom funded by

the National Science Foundation or elected to the National Academy of Sciences. They are outside of the scientific establishment and are kept out by those who regard themselves as real scientists.

If our only concern were to label certain people "pseudoscientists," we might simply check where their work is published and how their theories have been received by the scientific community. But we are concerned with the reasons certain doctrines are considered pseudoscientific; it is those reasons that interest philosophers of science.

Some philosophers have proposed necessary conditions for genuine science. That is, they have offered characteristics that any discipline or field of study must possess in order to qualify as genuine science. These characteristics are often called *demarcation criteria* because they can be used to differentiate science from its counterfeit: if a discipline fails to meet one of these conditions, then it is judged to be nonscientific.

In the twentieth century, philosophers of science have often disagreed about demarcation criteria. In this chapter Karl Popper, Thomas Kuhn, Imre Lakatos, and Paul Thagard each defend a different set of necessary conditions for genuine science. Popper's view, that a scientific theory must be open to refutation by making testable predictions, has been very influential, especially among working scientists. Kuhn, Lakatos, and Thagard all reject Popper's claim that falsifiability is the hallmark of genuine science but disagree about what should replace it. All three address whether a theory or discipline's claim to scientific legitimacy depends on historical considerations, such as how theories have developed over time.

The chapter ends with an exchange of views between Michael Ruse and Larry Laudan about the credentials of creation-science. Ruse, a prominent philosopher of biology, served as an expert witness in a trial concerning the constitutionality of an Arkansas law requiring public school biology teachers to present creationism as a viable scientific alternative to evolutionary theory. Under Ruse's guidance, the judge in the case drew up a list of five criteria for genuine science and concluded that creation-science failed on all five counts. Laudan not only criticizes the items on this list (which includes Popper's falsifiability) but also doubts whether there are any demarcation criteria that all scientific theories must satisfy.

■ | Notes

1. Quoted in Jack W. Grove, *In Defence of Science* (Toronto: University of Toronto Press, 1989), 137. See also Martin Gardner, *Science: Good, Bad and Bogus* (Buffalo, N.Y.: Prometheus Books, 1981), 185–206. The Parapsychological Association is still a member of the AAAS.

KARL POPPER

Science: Conjectures and Refutations

Mr. Turnbull had predicted evil consequences, . . . and was now doing the best in his power to bring about the verification of his own prophecies.

—Anthony Trollope

■ | I

When I received the list of participants in this course and realized that I had been asked to speak to philosophical colleagues* I thought, after some hesitation and consultation, that you would probably prefer me to speak about those problems which interest me most, and about those developments with which I am most intimately acquainted. I therefore decided to do what I have never done before: to give you a report on my own work in the philosophy of science, since the autumn of 1919 when I first began to grapple with the problem, 'When should a theory be ranked as scientific?' or 'Is there a criterion for the scientific character or status of a theory?'

The problem which troubled me at the time was neither, 'When is a theory true?' nor, 'When is a theory acceptable?' My problem was different. I wished to distinguish between science and pseudo-science; knowing very well that science often errs, and that pseudo-science may happen to stumble on the truth.

I knew, of course, the most widely accepted answer to my problem:

FROM Karl Popper, *Conjectures and Refutations* (London: Routledge and Kegan Paul, 1963), 33–39.

* This essay was originally presented as a lecture at Peterhouse College at Cambridge University in the summer of 1953 as part of a course on developments and trends in contemporary British philosophy, organized by the British Council. It was originally published as "Philosophy of Science: A Personal Report," in *British Philosophy in Mid-Century*, ed. C. A. Mace, (London: Allen and Unwin, 1957).

that science is distinguished from pseudo-science—or from ‘metaphysics’—by its *empirical method*, which is essentially *inductive*, proceeding from observation or experiment. But this did not satisfy me. On the contrary, I often formulated my problem as one of distinguishing between a genuinely empirical method and a non-empirical or even a pseudo-empirical method—that is to say, a method which, although it appeals to observation and experiment, nevertheless does not come up to scientific standards. The latter method may be exemplified by astrology, with its stupendous mass of empirical evidence based on observation—on horoscopes and on biographies.

But as it was not the example of astrology which led me to my problem I should perhaps briefly describe the atmosphere in which my problem arose and the examples by which it was stimulated. After the collapse of the Austrian Empire there had been a revolution in Austria: the air was full of revolutionary slogans and ideas, and new and often wild theories. Among the theories which interested me Einstein’s theory of relativity was no doubt by far the most important. Three others were Marx’s theory of history, Freud’s psycho-analysis, and Alfred Adler’s so-called ‘individual psychology’.*

There was a lot of popular nonsense talked about these theories, and especially about relativity (as still happens even today), but I was fortunate in those who introduced me to the study of this theory. We all—the small circle of students to which I belonged—were thrilled with the result of Eddington’s eclipse observations which in 1919 brought the first important confirmation of Einstein’s theory of gravitation. It was a great experience for us, and one which had a lasting influence on my intellectual development.†

The three other theories I have mentioned were also widely discussed among students at that time. I myself happened to come into personal contact with Alfred Adler, and even to co-operate with him in his social

* For a fascinating autobiographical account of Popper’s youthful flirtation and painful disenchantment with Marxism, see “A Crucial Year: Marxism, Science and Pseudoscience,” in *The Philosophy of Karl Popper*, ed. Paul A. Schilpp (La Salle, Ill.: Open Court, 1974), 1:23–29. There is also an extended criticism of Freud in Karl R. Popper, *Realism and the Aim of Science* (New York: Routledge, 1983), 163–74.

† Einstein’s general theory of relativity entails that light rays must bend in a gravitational field. Organized by Sir Arthur Eddington, two Royal Astronomical Society expeditions were dispatched to observe the solar eclipse of 1919, and verified that starlight was indeed deflected by the sun by the amount that Einstein had predicted. The *Times* of London reported this success as the most remarkable scientific event since the discovery of the planet Neptune. The light-bending test of relativity theory is discussed in “Popper’s Demarcation Criterion,” in the commentary on chapter 1, and in “Two Arguments for Explanationism,” in the commentary on chapter 4.

work among the children and young people in the working-class districts of Vienna where he had established social guidance clinics.

It was during the summer of 1919 that I began to feel more and more dissatisfied with these three theories—the Marxist theory of history, psycho-analysis, and individual psychology; and I began to feel dubious about their claims to scientific status. My problem perhaps first took the simple form, 'What is wrong with Marxism, psycho-analysis, and individual psychology? Why are they so different from physical theories, from Newton's theory, and especially from the theory of relativity?'

To make this contrast clear I should explain that few of us at the time would have said that we believed in the *truth* of Einstein's theory of gravitation. This shows that it was not my doubting the *truth* of those other three theories which bothered me, but something else. Yet neither was it that I merely felt mathematical physics to be more *exact* than the sociological or psychological type of theory. Thus what worried me was neither the problem of truth, at that stage at least, nor the problem of exactness or measurability. It was rather that I felt that these other three theories, though posing as sciences, had in fact more in common with primitive myths than with science; that they resembled astrology rather than astronomy.

I found that those of my friends who were admirers of Marx, Freud, and Adler, were impressed by a number of points common to these theories, and especially by their apparent *explanatory power*. These theories appeared to be able to explain practically everything that happened within the fields to which they referred. The study of any of them seemed to have the effect of an intellectual conversion or revelation, opening your eyes to a new truth hidden from those not yet initiated. Once your eyes were thus opened you saw confirming instances everywhere: the world was full of *verifications* of the theory. Whatever happened always confirmed it. Thus its truth appeared manifest; and unbelievers were clearly people who did not want to see the manifest truth; who refused to see it, either because it was against their class interest, or because of their repressions which were still 'un-analysed' and crying aloud for treatment.

The most characteristic element in this situation seemed to me the incessant stream of confirmations, of observations which 'verified' the theories in question; and this point was constantly emphasized by their adherents. A Marxist could not open a newspaper without finding on every page confirming evidence for his interpretation of history; not only in the news, but also in its presentation—which revealed the class bias of the paper—and especially of course in what the paper did *not* say. The Freudian analysts emphasized that their theories were constantly verified by their 'clinical observations'. As for Adler, I was much impressed by a personal experience. Once, in 1919, I reported to him a case which to me did not seem particularly Adlerian, but which he found no difficulty in analysing in terms of his theory of inferiority feelings, although he had not even

seen the child. Slightly shocked, I asked him how he could be so sure. 'Because of my thousandfold experience,' he replied; whereupon I could not help saying: 'And with this new case, I suppose, your experience has become thousand-and-one-fold.'

What I had in mind was that his previous observations may not have been much sounder than this new one; that each in its turn had been interpreted in the light of 'previous experience', and at the same time counted as additional confirmation. What, I asked myself, did it confirm? No more than that a case could be interpreted in the light of the theory. But this meant very little, I reflected, since every conceivable case could be interpreted in the light of Adler's theory, or equally of Freud's. I may illustrate this by two very different examples of human behaviour: that of a man who pushes a child into the water with the intention of drowning it; and that of a man who sacrifices his life in an attempt to save the child. Each of these two cases can be explained with equal ease in Freudian and in Adlerian terms. According to Freud the first man suffered from repression (say, of some component of his Oedipus complex), while the second man had achieved sublimation. According to Adler the first man suffered from feelings of inferiority (producing perhaps the need to prove to himself that he dared to commit some crime), and so did the second man (whose need was to prove to himself that he dared to rescue the child). I could not think of any human behaviour which could not be interpreted in terms of either theory. It was precisely this fact—that they always fitted, that they were always confirmed—which in the eyes of their admirers constituted the strongest argument in favour of these theories. It began to dawn on me that this apparent strength was in fact their weakness.

With Einstein's theory the situation was strikingly different. Take one typical instance—Einstein's prediction, just then confirmed by the findings of Eddington's expedition. Einstein's gravitational theory had led to the result that light must be attracted by heavy bodies (such as the sun), precisely as material bodies were attracted. As a consequence it could be calculated that light from a distant fixed star whose apparent position was close to the sun would reach the earth from such a direction that the star would seem to be slightly shifted away from the sun; or, in other words, that stars close to the sun would look as if they had moved a little away from the sun, and from one another. This is a thing which cannot normally be observed since such stars are rendered invisible in daytime by the sun's overwhelming brightness; but during an eclipse it is possible to take photographs of them. If the same constellation is photographed at night one can measure the distances on the two photographs, and check the predicted effect.

Now the impressive thing about this case is the *risk* involved in a prediction of this kind. If observation shows that the predicted effect is definitely absent, then the theory is simply refuted. The theory is *incompatible with certain possible results of observation*—in fact with results

which everybody before Einstein would have expected.¹ This is quite different from the situation I have previously described, when it turned out that the theories in question were compatible with the most divergent human behaviour, so that it was practically impossible to describe any human behaviour that might not be claimed to be a verification of these theories.

These considerations led me in the winter of 1919–20 to conclusions which I may now reformulate as follows.

- 1 It is easy to obtain confirmations, or verifications, for nearly every theory—if we look for confirmations.
- 2 Confirmations should count only if they are the result of *risky predictions*; that is to say, if, unenlightened by the theory in question, we should have expected an event which was incompatible with the theory—an event which would have refuted the theory.
- 3 Every ‘good’ scientific theory is a prohibition: it forbids certain things to happen. The more a theory forbids, the better it is.
- 4 A theory which is not refutable by any conceivable event is non-scientific. Irrefutability is not a virtue of a theory (as people often think) but a vice.
- 5 Every genuine test of a theory is an attempt to falsify it, or to refute it. Testability is falsifiability; but there are degrees of testability: some theories are more testable, more exposed to refutation, than others; they take, as it were, greater risks.
- 6 Confirming evidence should not count *except when it is the result of a genuine test of the theory*; and this means that it can be presented as a serious but unsuccessful attempt to falsify the theory. (I now speak in such cases of ‘corroborating evidence’.)
- 7 Some genuinely testable theories, when found to be false, are still upheld by their admirers—for example by introducing *ad hoc* some auxiliary assumption, or by re-interpreting the theory *ad hoc* in such a way that it escapes refutation. Such a procedure is always possible, but it rescues the theory from refutation only at the price of destroying, or at least lowering, its scientific status. (I later described such a rescuing operation as a ‘conventionalist twist’ or a ‘conventionalist stratagem’.)

One can sum up all this by saying that *the criterion of the scientific status of a theory is its falsifiability, or refutability, or testability.*

■ | II

I may perhaps exemplify this with the help of the various theories so far mentioned. Einstein’s theory of gravitation clearly satisfied the criterion of falsifiability. Even if our measuring instruments at the time did not allow

us to pronounce on the results of the tests with complete assurance, there was clearly a possibility of refuting the theory.

Astrology did not pass the test. Astrologers were greatly impressed, and misled, by what they believed to be confirming evidence—so much so that they were quite unimpressed by any unfavourable evidence. Moreover, by making their interpretations and prophecies sufficiently vague they were able to explain away anything that might have been a refutation of the theory had the theory and the prophecies been more precise. In order to escape falsification they destroyed the testability of their theory. It is a typical soothsayer's trick to predict things so vaguely that the predictions can hardly fail: that they become irrefutable.

The Marxist theory of history, in spite of the serious efforts of some of its founders and followers, ultimately adopted this soothsaying practice. In some of its earlier formulations (for example in Marx's analysis of the character of the 'coming social revolution') their predictions were testable, and in fact falsified.² Yet instead of accepting the refutations the followers of Marx re-interpreted both the theory and the evidence in order to make them agree. In this way they rescued the theory from refutation; but they did so at the price of adopting a device which made it irrefutable. They thus gave a 'conventionalist twist' to the theory; and by this stratagem they destroyed its much advertised claim to scientific status.

The two psycho-analytic theories were in a different class. They were simply non-testable, irrefutable. There was no conceivable human behaviour which could contradict them. This does not mean that Freud and Adler were not seeing certain things correctly: I personally do not doubt that much of what they say is of considerable importance, and may well play its part one day in a psychological science which is testable. But it does mean that those 'clinical observations' which analysts naïvely believe confirm their theory cannot do this any more than the daily confirmations which astrologers find in their practice.³ And as for Freud's epic of the Ego, the Super-ego, and the Id, no substantially stronger claim to scientific status can be made for it than for Homer's collected stories from Olympus. These theories describe some facts, but in the manner of myths. They contain most interesting psychological suggestions, but not in a testable form.

At the same time I realized that such myths may be developed, and become testable; that historically speaking all—or very nearly all—scientific theories originate from myths, and that a myth may contain important anticipations of scientific theories. Examples are Empedocles' theory of evolution by trial and error, or Parmenides' myth of the unchanging block universe in which nothing ever happens and which, if we add another dimension, becomes Einstein's block universe (in which, too, nothing ever happens, since everything is, four-dimensionally speaking, determined and laid down from the beginning). I thus felt that if a theory is found to be non-scientific, or 'metaphysical' (as we might say), it is not thereby found

to be unimportant, or insignificant, or 'meaningless', or 'nonsensical'.¹ But it cannot claim to be backed by empirical evidence in the scientific sense—although it may easily be, in some genetic sense, the 'result of observation'.

(There were a great many other theories of this pre-scientific or pseudo-scientific character, some of them, unfortunately, as influential as the Marxist interpretation of history; for example, the racialist interpretation of history—another of those impressive and all-explanatory theories which act upon weak minds like revelations.)

Thus the problem which I tried to solve by proposing the criterion of falsifiability was neither a problem of meaningfulness or significance, nor a problem of truth or acceptability. It was the problem of drawing a line (as well as this can be done) between the statements, or systems of statements, of the empirical sciences, and all other statements—whether they are of a religious or of a metaphysical character, or simply pseudo-scientific. Years later—it must have been in 1928 or 1929—I called this first problem of mine the '*problem of demarcation*'. The criterion of falsifiability is a solution to this problem of demarcation, for it says that statements or systems of statements, in order to be ranked as scientific, must be capable of conflicting with possible, or conceivable, observations. . . .

■ | Notes

1. This is a slight oversimplification, for about half of the Einstein effect may be derived from the classical theory, provided we assume a ballistic theory of light.
2. See, for example, my *Open Society and Its Enemies* [Routledge & Kegan Paul, 1945], ch. 15, section iii, and notes 13–14.
3. 'Clinical observations', like all other observations, are *interpretations in the light of theories* . . . ; and for this reason alone they are apt to seem to support those theories in the light of which they were interpreted. But real support can be obtained only from observations undertaken as tests (by 'attempted refutations'); and for this purpose *criteria of refutation* have to be laid down beforehand: it must be agreed which observable situations, if actually observed, mean that the theory is *refuted*. But what kind of clinical responses would refute to the satisfaction of the analyst not merely a particular analytic diagnosis but psycho-analysis itself? And have such criteria ever been discussed or agreed upon by analysts? Is there not, on the contrary, a whole family of analytic concepts, such as 'ambivalence' (I do not suggest that there is no such thing as ambivalence), which would make it difficult, if not impossible, to agree upon such criteria? Moreover, how much headway has been made in investigating the question of the extent to which the (conscious or unconscious) expectations and theories held by the analyst influence the 'clinical responses' of the patient? (To say nothing about the conscious attempts to influence the patient by proposing interpretations to him, etc.) Years ago I introduced the term '*Oedipus effect*' to describe the influence of a theory or expectation or prediction upon the event which it predicts or describes: it will be

remembered that the causal chain leading to Oedipus' parricide was started by the oracle's prediction of this event. This is a characteristic and recurrent theme of such myths, but one which seems to have failed to attract the interest of the analysts, perhaps not accidentally. (The problem of confirmatory dreams suggested by the analyst is discussed by Freud, for example in *Gesammelte Schriften* [Complete works], III, 1925, where he says on p. 314: 'If anybody asserts that most of the dreams which can be utilized in an analysis . . . owe their origin to [the analyst's] suggestion, then no objection can be made from the point of view of analytic theory. Yet there is nothing in this fact', he surprisingly adds, 'which would detract from the reliability of our results.')

4. The case of astrology, nowadays a typical pseudo-science, may illustrate this point. It was attacked, by Aristotelians and other rationalists, down to Newton's day, for the wrong reason—for its now accepted assertion that the planets had an 'influence' upon terrestrial ('sublunar') events. In fact Newton's theory of gravity, and especially the lunar theory of the tides, was historically speaking an offspring of astrological lore. Newton, it seems, was most reluctant to adopt a theory which came from the same stable as for example the theory that 'influenza' epidemics are due to an astral 'influence'. And Galileo, no doubt for the same reason, actually rejected the lunar theory of the tides; and his misgivings about Kepler may easily be explained by his misgivings about astrology.

Logic of Discovery or Psychology of Research?

Among the most fundamental issues on which Sir Karl [Popper] and I agree is our insistence that an analysis of the development of scientific knowledge must take account of the way science has actually been practiced. That being so, a few of his recurrent generalizations startle me. One of these provides the opening sentences of the first chapter of the *Logic of Scientific Discovery*: 'A scientist', writes Sir Karl, 'whether theorist or experimenter, puts forward statements, or systems of statements, and tests them step by step. In the field of the empirical sciences, more particularly, he constructs hypotheses, or systems of theories, and tests them against experience by observation and experiment.'¹ The statement is virtually a cliché, yet in application it presents three problems. It is ambiguous in its failure to specify which of two sorts of 'statements' or 'theories' are being tested. That ambiguity can, it is true, be eliminated by reference to other passages in Sir Karl's writings, but the generalization that results is historically mistaken. Furthermore, the mistake proves important, for the unambiguous form of the description misses just that characteristic of scientific practice which most nearly distinguishes the sciences from other creative pursuits.

There is one sort of 'statement' or 'hypothesis' that scientists do repeatedly subject to systematic test. I have in mind ~~statements~~ of an individual's best guesses about the proper way to connect his own research problem with the corpus of accepted scientific knowledge. He may, for example, conjecture that a given chemical unknown contains the salt of a rare earth, that the obesity of his experimental rats is due to a specified component in their diet, or that a newly discovered spectral pattern is to be understood as an effect of nuclear spin. In each case, the next steps in his research are intended to try out or test the conjecture or hypothesis.

FROM Imre Lakatos and Alan Musgrave, eds., *Criticism and the Growth of Knowledge* (Cambridge: Cambridge University Press, 1970) 4-10.

If it passes enough or stringent enough tests, the scientist has made a discovery or has at least resolved the puzzle he had been set. If not, he must either abandon the puzzle entirely or attempt to solve it with the aid of some other hypothesis. Many research problems, though by no means all, take this form. Tests of this sort are a standard component of what I have elsewhere labelled 'normal science' or 'normal research', an enterprise which accounts for the overwhelming majority of the work done in basic science. In no usual sense, however, are such tests directed to current theory. On the contrary, when engaged with a normal research problem, the scientist must *premise* current theory as the rules of his game. His object is to solve a puzzle, preferably one at which others have failed, and current theory is required to define that puzzle and to guarantee that, given sufficient brilliance, it can be solved.² Of course the practitioner of such an enterprise must often test the conjectural puzzle solution that his ingenuity suggests. But only his personal conjecture is tested. If it fails the test, only his own ability not the corpus of current science is impugned. In short, though tests occur frequently in normal science, these tests are of a peculiar sort, for in the final analysis it is the individual scientist rather than current theory which is tested.

This is not, however, the sort of test Sir Karl has in mind. He is above all concerned with the procedures through which science grows, and he is convinced that 'growth' occurs not primarily by accretion but by the revolutionary overthrow of an accepted theory and its replacement by a better one.³ (The subsumption under 'growth' of 'repeated overthrow' is itself a linguistic oddity whose *raison d'être* may become more visible as we proceed.) Taking this view, the tests which Sir Karl emphasizes are those which were performed to explore the limitations of accepted theory or to subject a current theory to maximum strain. Among his favourite examples, all of them startling and destructive in their outcome, are Lavoisier's experiments on calcination,* the eclipse expedition

* Calcination occurs when a metal is burned in air, forming a calx or oxide. According to the phlogiston theory, metals (and all other combustible substances) are compounds of an earthy calx and the fiery element, phlogiston. When a metal burns, the phlogiston is released, leaving the calx as a residue. Because metals gain weight when they are calcined, some proponents of the phlogiston theory conjectured that phlogiston must have negative weight. Others inferred that some other substance must combine with the metal when the phlogiston is released. By careful experiments in the 1770s, Antoine Lavoisier (1743-94) showed that the weight gained during calcination is entirely due to the metal combining with a gas in the air, which he named oxygen. Lavoisier's oxygen theory of calcination (and, more generally, of combustion) overthrew the phlogiston theory and gave rise to a revolution in chemistry. See James B. Conant, ed., *The Overthrow of the Phlogiston Theory: The Chemical Revolution of 1775-1789* (Cambridge, Mass.: Harvard University Press, 1950); reprinted in *Harvard Case Histories in Experimental Science*, ed. J. B. Conant and L. K. Nash (Cambridge, Mass.: Harvard University Press, 1966). See also Alan Musgrave, "Why Did Oxygen Supplant

of 1919,* and the recent experiments on parity conservation.†⁴ All, of course, are classic tests, but in using them to characterize scientific activity Sir Karl misses something terribly important about them. Episodes like these are very rare in the development of science. When they occur, they are generally called forth either by a prior crisis in the relevant field (Lavoisier's experiments or Lee and Yang's⁵) or by the existence of a theory which competes with the existing canons of research (Einstein's general relativity). These are, however, aspects of or occasions for what I have elsewhere called 'extraordinary research', an enterprise in which scientists do display very many of the characteristics Sir Karl emphasizes, but one which, at least in the past, has arisen only intermittently and under quite special circumstances in any scientific speciality.⁶

I suggest then that Sir Karl has characterized the entire scientific enterprise in terms that apply only to its occasional revolutionary parts. His emphasis is natural and common: the exploits of a Copernicus or Einstein make better reading than those of a Brahe or Lorentz;‡ Sir Karl

Phlogiston? Research Programmes in the Chemical Revolution," in *Method and Appraisal in the Physical Sciences*, ed. C. Howson (Cambridge: Cambridge University Press, 1976), 181-209.

* For information about the eclipse expedition of 1919 and its role in confirming Einstein's general theory of relativity, see the preceding reading by Karl Popper, "Science: Conjectures and Refutations." Further discussion can be found in "Popper's Demarcation Criterion," in the commentary on chapter 1, and in "Two Arguments for Explanationism," in the commentary on chapter 4.

† Kuhn is referring to the experiments performed by Chien-Shiung Wu and her associates in 1956-57, which verified the conjecture of Tsung Dao Lee and Chen Ning Yang that parity is not conserved in weak interactions. Wu's results were soon confirmed by other groups and Lee and Yang received the Nobel prize in physics in 1957 for their discovery of parity violation. For a description of Wu's experiment and an explanation of its revolutionary significance, see Eugene Wigner, "Violations of Symmetry in Physics," *Scientific American* 213 (1965): 28-36 and Martin Gardner, *The New Ambidextrous Universe*, 3d rev. ed. (New York: W. H. Freeman, 1990).

‡ For Kuhn, Tycho Brahe (1546-1601) and H. A. Lorentz (1853-1928) exemplify the conservative scientist practicing normal science. Brahe objected to Copernicus's revolutionary theory of a heliocentric universe on physical, astronomical, and religious grounds, proposing in its place his own version of a geostatic system. Like Ptolemy, Brahe had the sun moving around the earth, but unlike Ptolemy, he made the other planets orbit round the sun. In this way, Brahe was able to capture many of the explanatory features of Copernicus's theory without having to attribute any motion to the earth. Lorentz, like most physicists of his day, believed that light and other electromagnetic radiation propagates in an aether that is at rest with respect to absolute space. In order to account for the null result of the Michelson-Morley experiment, Lorentz (and, independently, Fitzgerald) postulated the famous Lorentz-Fitzgerald contraction according to which all physical objects contract in their direction of motion. Lorentz later introduced time dilation, thus obtaining the Lorentz transformations that lie at the heart of Einstein's special

would not be the first if he mistook what I call normal science for an intrinsically uninteresting enterprise. Nevertheless, neither science nor the development of knowledge is likely to be understood if research is viewed exclusively through the revolutions it occasionally produces. For example, though testing of basic commitments occurs only in extraordinary science, it is normal science that discloses both the points to test and the manner of testing. Or again, it is for the normal, not the extraordinary practice of science that professionals are trained; if they are nevertheless eminently successful in displacing and replacing the theories on which normal practice depends, that is an oddity which must be explained. Finally, and this is for now my main point, a careful look at the scientific enterprise suggests that it is normal science, in which Sir Karl's sort of testing does not occur, rather than extraordinary science which most nearly distinguishes science from other enterprises. If a demarcation criterion exists (we must not, I think, seek a sharp or decisive one), it may lie just in that part of science which Sir Karl ignores.

In one of his most evocative essays, Sir Karl traces the origin of 'the tradition of critical discussion [which] represents the only practicable way of expanding our knowledge' to the Greek philosophers between Thales and Plato, the men who, as he sees it, encouraged critical discussion both between schools and within individual schools.⁷ The accompanying description of Presocratic discourse is most apt, but what is described does not at all resemble science. Rather it is the tradition of claims, counter-claims, and debates over fundamentals which, except perhaps during the Middle Ages, have characterized philosophy and much of social science ever since. Already by the Hellenistic period mathematics, astronomy, statics and the geometric parts of optics had abandoned this mode of discourse in favour of puzzle solving. Other sciences, in increasing numbers, have undergone the same transition since. In a sense, to turn Sir Karl's view on its head, it is precisely the abandonment of critical discourse that marks the transition to a science. Once a field has made that transition, critical discourse recurs only at moments of crisis when the bases of the field are again in jeopardy.⁸ Only when they must choose between competing theories do scientists behave like philosophers. That, I think, is why Sir Karl's brilliant description of the reasons for the choice between metaphysical systems so closely resembles my description of the reasons for choosing between scientific theories.⁹ In neither choice, as I shall shortly try to show, can testing play a quite decisive role.

theory of relativity; but unlike Einstein, Lorentz worked within a classical framework of absolute space and time. For introductory accounts of the contrast between the theories of Lorentz and Einstein and their differing interpretations of the Michelson-Morley experiment, see Albert Einstein, *Relativity: The Special and General Theory*, trans. R. W. Lawson (New York: Crown, 1961), and Jonathan Powers, *Philosophy and the New Physics* (New York: Methuen, 1982), ch. 3.

There is, however, good reason why testing has seemed to do so, and in exploring it Sir Karl's duck may at last become my rabbit.* No puzzle-solving enterprise can exist unless its practitioners share criteria which, for that group and for that time, determine when a particular puzzle has been solved. The same criteria necessarily determine failure to achieve a solution, and anyone who chooses may view that failure as the failure of a theory to pass a test. Normally, as I have already insisted, it is not viewed that way. Only the practitioner is blamed, not his tools. But under the special circumstances which induce a crisis in the profession (e.g. gross failure, or repeated failure by the most brilliant professionals) the group's opinion may change. A failure that had previously been personal may then come to seem the failure of a theory under test. Thereafter, because the test arose from a puzzle and thus carried settled criteria of solution, it proves both more severe and harder to evade than the tests available within a tradition whose normal mode is critical discourse rather than puzzle solving.

In a sense, therefore, severity of test-criteria is simply one side of the coin whose other face is a puzzle-solving tradition. That is why Sir Karl's line of demarcation and my own so frequently coincide. That coincidence is, however, only in their *outcome*; the *process* of applying them is very different, and it isolates distinct aspects of the activity about which the decision—science or non-science—is to be made. Examining the vexing cases, for example, psychoanalysis or Marxist historiography, for which Sir Karl tells us his criterion was initially designed,¹⁰ I concur that they cannot now properly be labelled 'science'. But I reach that conclusion by a route far surer and more direct than his. One brief example may suggest that of the two criteria, testing and puzzle solving, the latter is at once the less equivocal and the more fundamental.

To avoid irrelevant contemporary controversies, I consider astrology rather than, say, psychoanalysis. Astrology is Sir Karl's most frequently cited example of a 'pseudo-science'.¹¹ He says: 'By making their interpretations and prophecies sufficiently vague they [astrologers] were able to explain away anything that might have been a refutation of the theory had the theory and the prophecies been more precise. In order to escape falsification they destroyed the testability of their theory.'¹² Those generalizations catch something of the spirit of the astrological enterprise. But taken at all literally, as they must be if they are to provide a demarcation criterion, they are impossible to support. The history of astrology during the cen-

* The duck-rabbit is a visually ambiguous drawing, made popular among philosophers by Ludwig Wittgenstein in his *Philosophical Investigations* (1953). It can be seen either as a duck's head with a long beak or as a rabbit's head with long ears, but it cannot be seen as both at the same time. It is a favorite with philosophers of science (such as Kuhn, Hanson, and Feyerabend) wishing to emphasize the theory-ladenness of observation.

turies when it was intellectually reputable records many predictions that categorically failed.¹³ Not even astrology's most convinced and vehement exponents doubted the recurrence of such failures. Astrology cannot be barred from the sciences because of the form in which its predictions were cast.

Nor can it be barred because of the way its practitioners explained failure. Astrologers pointed out, for example, that, unlike general predictions about, say, an individual's propensities or a natural calamity, the forecast of an individual's future was an immensely complex task, demanding the utmost skill, and extremely sensitive to minor errors in relevant data. The configuration of the stars and eight planets was constantly changing; the astronomical tables used to compute the configuration at an individual's birth were notoriously imperfect; few men knew the instant of their birth with the requisite precision.¹⁴ No wonder, then, that forecasts often failed. Only after astrology itself became implausible did these arguments come to seem question-begging.¹⁵ Similar arguments are regularly used today when explaining, for example, failures in medicine or meteorology. In times of trouble they are also deployed in the exact sciences, fields like physics, chemistry, and astronomy.¹⁶ There was nothing unscientific about the astrologer's explanation of failure.

Nevertheless, astrology was not a science. Instead it was a craft, one of the practical arts, with close resemblances to engineering, meteorology, and medicine as these fields were practised until little more than a century ago. The parallels to an older medicine and to contemporary psychoanalysis are, I think, particularly close. In each of these fields shared theory was adequate only to establish the plausibility of the discipline and to provide a rationale for the various craft-rules which governed practice. These rules had proved their use in the past, but no practitioner supposed they were sufficient to prevent recurrent failure. A more articulated theory and more powerful rules were desired, but it would have been absurd to abandon a plausible and badly needed discipline with a tradition of limited success simply because these desiderata were not yet at hand. In their absence, however, neither the astrologer nor the doctor could do research. Though they had rules to apply, they had no puzzles to solve and therefore no science to practise.¹⁷

Compare the situations of the astronomer and the astrologer. If an astronomer's prediction failed and his calculations checked, he could hope to set the situation right. Perhaps the data were at fault: old observations could be re-examined and new measurements made, tasks which posed a host of calculational and instrumental puzzles. Or perhaps theory needed adjustment, either by the manipulation of epicycles, eccentrics, equants, etc., or by more fundamental reforms of astronomical technique. For more than a millennium these were the theoretical and mathematical puzzles around which, together with their instrumental counterparts, the astronomical research tradition was constituted. The astrologer, by contrast, had

no such puzzles. The occurrence of failures could be explained, but particular failures did not give rise to research puzzles, for no man, however skilled, could make use of them in a constructive attempt to revise the astrological tradition. There were too many possible sources of difficulty, most of them beyond the astrologer's knowledge, control, or responsibility. Individual failures were correspondingly uninformative, and they did not reflect on the competence of the prognosticator in the eyes of his professional compeers.¹⁸ Though astronomy and astrology were regularly practised by the same people, including Ptolemy, Kepler, and Tycho Brahe, there was never an astrological equivalent of the puzzle-solving astronomical tradition. And without puzzles, able first to challenge and then to attest the ingenuity of the individual practitioner, astrology could not have become a science even if the stars had, in fact, controlled human destiny.

In short, though astrologers made testable predictions and recognized that these predictions sometimes failed, they did not and could not engage in the sorts of activities that normally characterize all recognized sciences. Sir Karl is right to exclude astrology from the sciences, but his over-concentration on science's occasional revolutions prevents his seeing the surest reason for doing so.

That fact, in turn, may explain another oddity of Sir Karl's historiography. Though he repeatedly underlines the role of tests in the replacement of scientific theories, he is also constrained to recognize that many theories, for example the Ptolemaic, were replaced before they had in fact been tested.¹⁹ On some occasions, at least, tests are not requisite to the revolutions through which science advances. But that is not true of puzzles. Though the theories Sir Karl cites had not been put to the test before their displacement, none of these was replaced before it had ceased adequately to support a puzzle-solving tradition. The state of astronomy was a scandal in the early sixteenth century. Most astronomers nevertheless felt that normal adjustments of a basically Ptolemaic model would set the situation right. In this sense the theory had not failed a test. But a few astronomers, Copernicus among them, felt that the difficulties must lie in the Ptolemaic approach itself rather than in the particular versions of Ptolemaic theory so far developed, and the results of that conviction are already recorded. The situation is typical.²⁰ With or without tests, a puzzle-solving tradition can prepare the way for its own displacement. To rely on testing as the mark of a science is to miss what scientists mostly do and, with it, the most characteristic feature of their enterprise. . . .

■ | Notes

1. Popper [1959], p. 27.

2. For an extended discussion of normal science, the activity which practitioners are trained to carry on, see my [1962], pp. 23-42, and 135-42. It is important to

notice that when I describe the scientist as a puzzle solver and Sir Karl describes him as a problem solver (e.g. in his [1963], pp. 67, 222), the similarity of our terms disguises a fundamental divergence. Sir Karl writes (the italics are his), 'Admittedly, our expectations, and thus our theories, may precede, historically, even our problems. *Yet science starts only with problems.* Problems crop up especially when we are disappointed in our expectations, or when our theories involve us in difficulties, in contradictions'. I use the term 'puzzle' in order to emphasize that the difficulties which *ordinarily* confront even the very best scientists are, like crossword puzzles or chess puzzles, challenges only to his ingenuity. He is in difficulty, not current theory. My point is almost the converse of Sir Karl's.

3. Cf. Popper [1963], pp. 129, 215 and 221, for particularly forceful statements of this position.

4. For example, Popper [1963], p. 220.

5. For the work on calcination see, Guerlac [1961]. For the background of the parity experiments see, Hafner and Presswood [1965].

6. The point is argued at length in my [1962], pp. 52-97.

7. Popper [1963], chapter 5, especially pp. 148-52.

8. Though I was not then seeking a demarcation criterion, just these points are argued at length in my [1962], pp. 10-22 and 87-90.

9. Cf. Popper [1963], pp. 192-200, with my [1962], pp. 143-58.

10. Popper [1963], p. 34 [p. 4-5, above].

11. The index to Popper [1963] has eight entries under the heading 'astrology as a typical pseudo-science'.

12. Popper [1963], p. 37 [p. 8, above].

13. For examples see, Thorndike [1923-58], 5, pp. 225 ff.; 6, pp. 71, 101, 114.

14. For reiterated explanations of failure see, *ibid.* I, pp. 11 and 514 f.; 4, 368; 5, 279.

15. A perceptive account of some reasons for astrology's loss of plausibility is included in Stahlman [1956]. For an explanation of astrology's previous appeal see, Thorndike [1955].

16. Cf. my [1962], pp. 66-76.

17. This formulation suggests that Sir Karl's criterion of demarcation might be saved by a minor restatement entirely in keeping with his apparent intent. For a field to be a science its conclusions must be *logically derivable* from *shared premises*. On this view astrology is to be barred not because its forecasts were not testable but because only the most general and least testable ones could be derived from accepted theory. Since any field that did satisfy this condition *might* support a puzzle solving tradition, the suggestion is clearly helpful. It comes close to supplying a sufficient condition for a field's being a science. But in this form, at least, it is not even quite a sufficient condition, and it is surely not a necessary one. It would, for example, admit surveying and navigation as sciences, and it would bar taxonomy, historical geology, and the theory of evolution. The conclusions of a

science may be both precise and binding without being fully derivable by logic from accepted premises. Cf. my [1962], pp. 35–51. . . .

18. This is not to suggest that astrologers did not criticize each other. On the contrary, like practitioners of philosophy and some social sciences, they belonged to a variety of different schools, and the inter-school strife was sometimes bitter. But these debates ordinarily revolved about the *implausibility* of the particular theory employed by one or another school. Failures of individual predictions played very little role. Compare Thorndike [1923–58], 5, p. 233.

19. Cf. Popper [1963], p. 246.

20. Cf. my [1962], pp. 77–87.

■ | References

Braithwaite [1953]: *Scientific Explanation*, 1953.

Guerlac [1961]: *Lavoisier—The Crucial Year*, 1961.

Hafner and Presswood [1965]: 'Strong Interference and Weak Interactions', *Science*, 149, pp. 503–10.

Hawkins [1963]: Review of Kuhn's 'Structure of Scientific Revolutions', *American Journal of Physics*, 31.

Hempel [1965]: *Aspects of Scientific Explanation*, 1965.

Lakatos [1963–4]: 'Proofs and Refutations', *The British Journal for the Philosophy of Science*, 14, pp. 1–25, 120–39, 221–43, 296–342.

Kuhn [1958]: 'The Role of Measurement in the Development of Physical Science', *Isis*, 49, pp. 161–93.

Kuhn [1962]: *The Structure of Scientific Revolutions*, 1962.

Popper [1935]: *Logik der Forschung*, 1935.

Popper [1945]: *The Open Society and its Enemies*, 2 vols, 1945.

Popper [1957]: *The Poverty of Historicism*, 1957.

Popper [1959]: *Logic of Scientific Discovery*, 1959.

Popper [1963]: *Conjectures and Refutations*, 1963.

Stahlman [1956]: 'Astrology in Colonial America: An Extended Query', *William and Mary Quarterly*, 13, pp. 551–63.

Thorndike [1923–58]: *A History of Magic and Experimental Science*, 8 vols, 1923–58.

Thorndike [1955]: 'The True Place of Astrology in the History of Science', *Isis*, 46, pp. 273–8.

Science and Pseudoscience

Man's respect for knowledge is one of his most peculiar characteristics. Knowledge in Latin is *scientia*, and science came to be the name of the most respectable kind of knowledge. But what distinguishes knowledge from superstition, ideology or pseudoscience? The Catholic Church excommunicated Copernicans, the Communist Party persecuted Mendelians on the ground that their doctrines were pseudoscientific. The demarcation between science and pseudoscience is not merely a problem of armchair philosophy: it is of vital social and political relevance.

Many philosophers have tried to solve the problem of demarcation in the following terms: a statement constitutes knowledge if sufficiently many people believe it sufficiently strongly. But the history of thought shows us that many people were totally committed to absurd beliefs. If the strength of beliefs were a hallmark of knowledge, we should have to rank some tales about demons, angels, devils, and of heaven and hell as knowledge. Scientists, on the other hand, are very sceptical even of their best theories. Newton's is the most powerful theory science has yet produced, but Newton himself never believed that bodies attract each other at a distance. So no degree of commitment to beliefs makes them knowledge. Indeed, the hallmark of scientific behaviour is a certain scepticism even towards one's most cherished theories. Blind commitment to a theory is not an intellectual virtue: it is an intellectual crime.

Thus a statement may be pseudoscientific even if it is eminently 'plausible' and everybody believes in it, and it may be scientifically valuable even if it is unbelievable and nobody believes in it. A theory may even be of supreme scientific value even if no one understands it, let alone believes it.

FROM Imre Lakatos, *Philosophical Papers*, vol. 1 (Cambridge: Cambridge University Press, 1977), 1-7. Written in early 1973, this was originally presented as a radio lecture broadcast by the Open University (30 June 1973).

The cognitive value of a theory has nothing to do with its psychological influence on people's minds. Belief, commitment, understanding are states of the human mind. But the objective, scientific value of a theory is independent of the human mind which creates it or understands it. Its scientific value depends only on what objective support these conjectures have in facts. As Hume said:

If we take in our hand any volume; of divinity, or school metaphysics, for instance; let us ask, does it contain any abstract reasoning concerning quantity or number? No. Does it contain any experimental reasoning concerning matter of fact and existence? No. Commit it then to the flames. For it can contain nothing but sophistry and illusion.*

But what is 'experimental' reasoning? If we look at the vast seventeenth-century literature on witchcraft, it is full of reports of careful observations and sworn evidence—even of experiments. Glanvill, the house philosopher of the early Royal Society, regarded witchcraft as the paradigm of experimental reasoning. We have to define experimental reasoning before we start Humean book burning.

In scientific reasoning, theories are confronted with facts; and one of the central conditions of scientific reasoning is that theories must be supported by facts. Now how exactly can facts support theory?

Several different answers have been proposed. Newton himself thought that he proved his laws from facts. He was proud of not uttering mere hypotheses: he only published theories proven from facts. In particular, he claimed that he deduced his laws from the 'phenomena' provided by Kepler. But his boast was nonsense, since according to Kepler, planets move in ellipses, but according to Newton's theory, planets would move in ellipses only if the planets did not disturb each other in their motion. But they do. This is why Newton had to devise a perturbation theory from which it follows that no planet moves in an ellipse.

One can today easily demonstrate that there can be no valid derivation of a law of nature from any finite number of facts; but we still keep reading about scientific theories being proved from facts. Why this stubborn resistance to elementary logic?

There is a very plausible explanation. Scientists want to make their theories respectable, deserving of the title 'science', that is, genuine knowledge. Now the most relevant knowledge in the seventeenth century, when science was born, concerned God, the Devil, Heaven and Hell. If one got one's conjectures about matters of divinity wrong, the consequence of one's mistake was eternal damnation. Theological knowledge cannot be

* These famous lines are from the final paragraph of David Hume's *An Enquiry Concerning Human Understanding*, first published in 1748 (under the title *Philosophical Essays Concerning Human Understanding*).

fallible: it must be beyond doubt. Now the Enlightenment thought that we were fallible and ignorant about matters theological. There is no scientific theology and, therefore, no theological knowledge. Knowledge can only be about Nature, but this new type of knowledge had to be judged by the standards they took over straight from theology: it had to be proven beyond doubt. Science had to achieve the very certainty which had escaped theology. A scientist, worthy of the name, was not allowed to guess: he had to prove each sentence he uttered from facts. This was the criterion of scientific honesty. Theories unproven from facts were regarded as sinful pseudoscience, heresy in the scientific community.

It was only the downfall of Newtonian theory in this century which made scientists realize that their standards of honesty had been utopian. Before Einstein most scientists thought that Newton had deciphered God's ultimate laws by proving them from the facts. Ampère, in the early nineteenth century, felt he had to call his book on his speculations concerning electromagnetism: *Mathematical Theory of Electrodynamical Phenomena Unequivocally Deduced from Experiment*. But at the end of the volume he casually confesses that some of the experiments were never performed and even that the necessary instruments had not been constructed!

If all scientific theories are equally unprovable, what distinguishes scientific knowledge from ignorance, science from pseudoscience?

One answer to this question was provided in the twentieth century by 'inductive logicians'. Inductive logic set out to define the probabilities of different theories according to the available total evidence. If the mathematical probability of a theory is high, it qualifies as scientific; if it is low or even zero, it is not scientific. Thus the hallmark of scientific honesty would be never to say anything that is not at least highly probable. Probabilism has an attractive feature: instead of simply providing a black-and-white distinction between science and pseudoscience, it provides a continuous scale from poor theories with low probability to good theories with high probability. But, in 1934, Karl Popper, one of the most influential philosophers of our time, argued that the mathematical probability of all theories, scientific or pseudoscientific, given *any* amount of evidence is zero.* If Popper is right, scientific theories are not only equally unprovable but also equally improbable. A new demarcation criterion was needed and Popper proposed a rather stunning one. A theory may be scientific even if there is not a shred of evidence in its favour, and it may be pseudoscientific even if all the available evidence is in its favour. That is, the scientific or non-scientific character of a theory can be determined independently of the facts. A theory is 'scientific' if one is prepared to specify in advance a crucial experiment (or observation) which can falsify it, and it is pseudoscientific if one refuses to specify such a 'potential

* Popper's argument for this claim can be found in Appendix *vii of *The Logic of Scientific Discovery* (New York: Basic Books, 1959), 363-67.

falsifier'. But if so, we do not demarcate scientific theories from pseudoscientific ones, but rather ~~scientific method from non-scientific method~~. Marxism, for a Popperian, is scientific if the Marxists are prepared to specify facts which, if observed, make them give up Marxism. If they refuse to do so, Marxism becomes a pseudoscience. It is always interesting to ask a Marxist, what conceivable event would make him abandon his Marxism. If he is committed to Marxism, he is bound to find it immoral to specify a state of affairs which can falsify it. Thus a proposition may petrify into pseudoscientific dogma or become genuine knowledge, depending on whether we are prepared to state observable conditions which would refute it.

Is, then, Popper's falsifiability criterion the solution to the problem of demarcating science from pseudoscience? No. For Popper's criterion ignores the remarkable tenacity of scientific theories. Scientists have thick skins. They do not abandon a theory merely because facts contradict it. They normally either invent some rescue hypothesis to explain what they then call a mere anomaly or, if they cannot explain the anomaly, they ignore it, and direct their attention to other problems. Note that scientists talk about anomalies, recalcitrant instances, not refutations. History of science, of course, is full of accounts of how crucial experiments allegedly killed theories. But such accounts are fabricated long after the theory had been abandoned. Had Popper ever asked a Newtonian scientist under what experimental conditions he would abandon Newtonian theory, some Newtonian scientists would have been exactly as nonplussed as are some Marxists.

What, then, is the hallmark of science? Do we have to capitulate and agree that a scientific revolution is just an irrational change in commitment, that it is a religious conversion? Tom Kuhn, a distinguished American philosopher of science, arrived at this conclusion after discovering the naïvety of Popper's falsificationism. But if Kuhn is right, then there is no explicit demarcation between science and pseudoscience, no distinction between scientific progress and intellectual decay, there is no objective standard of honesty. But what criteria can he then offer to demarcate scientific progress from intellectual degeneration?

In the last few years I have been advocating a methodology of scientific research programmes, which solves some of the problems which both Popper and Kuhn failed to solve.

First, I claim that the typical descriptive unit of great scientific achievements is not an isolated hypothesis but rather a research programme. Science is not simply trial and error, a series of conjectures and refutations. 'All swans are white' may be falsified by the discovery of one black swan. But such trivial trial and error does not rank as science. Newtonian science, for instance, is not simply a set of four conjectures—the three laws of mechanics and the law of gravitation. These four laws constitute only the 'hard core' of the Newtonian programme. But this hard

core is tenaciously protected from refutation by a vast 'protective belt' of auxiliary hypotheses. And, even more importantly, the research programme also has a 'heuristic', that is, a powerful problem-solving machinery, which, with the help of sophisticated mathematical techniques, digests anomalies and even turns them into positive evidence. For instance, if a planet does not move exactly as it should, the Newtonian scientist checks his conjectures concerning atmospheric refraction, concerning propagation of light in magnetic storms, and hundreds of other conjectures which are all part of the programme. He may even invent a hitherto unknown planet and calculate its position, mass and velocity in order to explain the anomaly.

Now, Newton's theory of gravitation, Einstein's relativity theory, quantum mechanics, Marxism, Freudianism, are all research programmes, each with a characteristic hard core stubbornly defended, each with its more flexible protective belt and each with its elaborate problem-solving machinery. Each of them, at any stage of its development, has unsolved problems and undigested anomalies. All theories, in this sense, are born refuted and die refuted. But are they equally good? Until now I have been describing what research programmes are like. But how can one distinguish a scientific or progressive programme from a pseudoscientific or degenerating one?

Contrary to Popper, the difference cannot be that some are still unrefuted, while others are already refuted. When Newton published his *Principia*, it was common knowledge that it could not properly explain even the motion of the moon; in fact, lunar motion refuted Newton. Kaufmann, a distinguished physicist, refuted Einstein's relativity theory in the very year it was published.* But all the research programmes I admire have one characteristic in common. They all predict novel facts, facts which had been either undreamt of, or have indeed been contradicted by previous or rival programmes. In 1686, when Newton published his theory of gravitation, there were, for instance, two current theories concerning comets. The more popular one regarded comets as a signal from an angry God warning that He will strike and bring disaster. A little known theory of Kepler's held that comets were celestial bodies moving along straight lines. Now according to Newtonian theory, some of them moved in hy-

* Here, as elsewhere in this reading, Lakatos is using the word *refuted* rather loosely. For Lakatos, a refutation is any apparently well-founded result that seems to be inconsistent with a theory. In the two cases he mentions — Newton and the moon, Einstein and Kaufmann's experiments on beta rays — the "refutations" were later shown to be spurious: the moon's motion is not actually inconsistent with Newton's theory, and Kaufmann's results were due to experimental error. For an account of Kaufmann's experiments and Einstein's reaction to them, see Arthur I. Miller, *Albert Einstein's Special Theory of Relativity* (Reading, Mass.: Addison-Wesley, 1981).

perbolas or parabolas never to return; others moved in ordinary ellipses. Halley, working in Newton's programme, calculated on the basis of observing a brief stretch of a comet's path that it would return in seventy-two years' time; he calculated to the minute when it would be seen again at a well-defined point of the sky. This was incredible. But seventy-two years later, when both Newton and Halley were long dead, Halley's comet returned exactly as Halley predicted. Similarly, Newtonian scientists predicted the existence and exact motion of small planets which had never been observed before. Or let us take Einstein's programme. This programme made the stunning prediction that if one measures the distance between two stars in the night and if one measures the distance between them during the day (when they are visible during an eclipse of the sun), the two measurements will be different. Nobody had thought to make such an observation before Einstein's programme. Thus, in a progressive research programme, theory leads to the discovery of hitherto unknown novel facts. In degenerating programmes, however, theories are fabricated only in order to accommodate known facts. Has, for instance, Marxism ever predicted a stunning novel fact successfully? Never! It has some famous unsuccessful predictions. It predicted the absolute impoverishment of the working class. It predicted that the first socialist revolution would take place in the industrially most developed society. It predicted that socialist societies would be free of revolutions. It predicted that there will be no conflict of interests between socialist countries. Thus the early predictions of Marxism were bold and stunning but they failed. Marxists explained all their failures: they explained the rising living standards of the working class by devising a theory of imperialism; they even explained why the first socialist revolution occurred in industrially backward Russia. They 'explained' Berlin 1953, Budapest 1956, Prague 1968. They 'explained' the Russian-Chinese conflict. But their auxiliary hypotheses were all cooked up after the event to protect Marxian theory from the facts. The Newtonian programme led to novel facts; the Marxian lagged behind the facts and has been running fast to catch up with them.

To sum up. The hallmark of empirical progress is not trivial verifications: Popper is right that there are millions of them. It is no success for Newtonian theory that stones, when dropped, fall towards the earth, no matter how often this is repeated. But so-called 'refutations' are not the hallmark of empirical failure, as Popper has preached, since all programmes grow in a permanent ocean of anomalies. What really count are dramatic, unexpected, stunning predictions: a few of them are enough to tilt the balance; where theory lags behind the facts, we are dealing with miserable degenerating research programmes.

Now, how do scientific revolutions come about? If we have two rival research programmes, and one is progressing while the other is degenerating, scientists tend to join the progressive programme. This is the ra-

tionale of scientific revolutions. But while it is a matter of intellectual honesty to keep the record public, it is not dishonest to stick to a degenerating programme and try to turn it into a progressive one.

As opposed to Popper the methodology of scientific research programmes does not offer instant rationality. One must treat budding programmes leniently: programmes may take decades before they get off the ground and become empirically progressive. Criticism is not a Popperian quick kill, by refutation. Important criticism is always constructive: there is no refutation without a better theory. Kuhn is wrong in thinking that scientific revolutions are sudden, irrational changes in vision. The history of science refutes both Popper and Kuhn: on close inspection both Popperian crucial experiments and Kuhnian revolutions turn out to be myths: what normally happens is that progressive research programmes replace degenerating ones.

The problem of demarcation between science and pseudoscience has grave implications also for the institutionalization of criticism. Copernicus's theory was banned by the Catholic Church in 1616 because it was said to be pseudoscientific. It was taken off the index in 1820 because by that time the Church deemed that facts had proved it and therefore it became scientific. The Central Committee of the Soviet Communist Party in 1949 declared Mendelian genetics pseudoscientific and had its advocates, like Academician Vavilov, killed in concentration camps; after Vavilov's murder Mendelian genetics was rehabilitated; but the Party's right to decide what is science and publishable and what is pseudoscience and punishable was upheld. The new liberal Establishment of the West also exercises the right to deny freedom of speech to what it regards as pseudoscience, as we have seen in the case of the debate concerning race and intelligence. All these judgments were inevitably based on some sort of demarcation criterion. This is why the problem of demarcation between science and pseudoscience is not a pseudo-problem of armchair philosophers: it has grave ethical and political implications.

PAUL R. THAGARD

Why Astrology Is a Pseudoscience

Most philosophers and historians of science agree that astrology is a pseudoscience, but there is little agreement on *why* it is a pseudoscience. Answers range from matters of verifiability and falsifiability, to questions of progress and Kuhnian normal science, to the different sorts of objections raised by a large panel of scientists recently organized by *The Humanist* magazine. Of course there are also Feyerabendian anarchists* and others who say that no demarcation of science from pseudoscience is possible. However, I shall propose a complex criterion for distinguishing disciplines as pseudoscientific; this criterion is unlike verificationist and falsificationist attempts in that it introduces social and historical features as well as logical ones.

I begin with a brief description of astrology. It would be most unfair to evaluate astrology by reference to the daily horoscopes found in newspapers and popular magazines. These horoscopes deal only with sun signs, whereas a full horoscope makes reference to the "influences" also of the moon and the planets, while also discussing the ascendant sign and other matters.

Astrology divides the sky into twelve regions, represented by the familiar signs of the Zodiac: Aquarius, Libra and so on. The sun sign represents the part of the sky occupied by the sun at the time of birth. For example, anyone born between September 23 and October 22 is a Libran. The ascendant sign, often assumed to be at least as important as the sun

FROM P. Asquith and I. Hacking, eds., *Proceedings of the Philosophy of Science Association* Vol. 1 (East Lansing, Mich.: Philosophy of Science Association, 1978), 223-34.

* Paul Feyerabend (1924-94) used the term *epistemological anarchism* in his *Against Method* (London: New Left Books, 1975), arguing that there is no rational method in science and that the only principle consistent with scientific progress is "anything goes."

sign, represents the part of the sky rising on the eastern horizon at the time of birth, and therefore changes every two hours. To determine this sign, accurate knowledge of the time and place of birth is essential. The moon and the planets (of which there are five or eight depending on whether Uranus, Neptune and Pluto are taken into account) are also located by means of charts on one of the parts of the Zodiac. Each planet is said to exercise an influence in a special sphere of human activity; for example. Mars governs drive, courage and daring, while Venus governs love and artistic endeavor. The immense number of combinations of sun, ascendant, moon and planetary influences allegedly determines human personality, behavior and fate.

Astrology is an ancient practice, and appears to have its origins in Chaldea, thousands of years B.C. By 700 B.C., the Zodiac was established, and a few centuries later the signs of the Zodiac were very similar to current ones. The conquests of Alexander the Great brought astrology to Greece, and the Romans were exposed in turn. Astrology was very popular during the fall of the Republic, with many notables such as Julius Caesar having their horoscopes cast. However, there was opposition from such men as Lucretius and Cicero.

Astrology underwent a gradual codification culminating in Ptolemy's *Tetrabiblos* [20], written in the second century A.D. This work describes in great detail the powers of the sun, moon and planets, and their significance in people's lives. It is still recognized as a fundamental textbook of astrology. Ptolemy took astrology as seriously as he took his famous work in geography and astronomy; this is evident from the introduction to the *Tetrabiblos*, where he discusses two available means of making predictions based on the heavens. The first and admittedly more effective of these concerns the relative movements of the sun, moon and planets, which Ptolemy had already treated in his celebrated *Almagest* [19]. The secondary but still legitimate means of prediction is that in which we use the "natural character" of the aspects of movement of heavenly bodies to "investigate the changes which they bring about in that which they surround." ([20], p. 3). He argues that this method of prediction is possible because of the manifest effects of the sun, moon and planets on the earth, for example on weather and the tides.

The European Renaissance is heralded for the rise of modern science, but occult arts such as astrology and alchemy flourished as well. Arthur Koestler has described Kepler's interest in astrology: not only did astrology provide Kepler with a livelihood, he also pursued it as a serious interest, although he was skeptical of the particular analyses of previous astrologers ([13], pp. 244-248). Astrology was popular both among intellectuals and the general public through the seventeenth century. However, astrology lost most of this popularity in the eighteenth century, when it was attacked by such figures of the Enlightenment as Swift [24] and Voltaire [29]. Only since the 1930's has astrology again gained a huge audience: most people

today know at least their sun signs, and a great many believe that the stars and planets exercise an important influence on their lives.

In an attempt to reverse this trend, Bart Bok, Lawrence Jerome and Paul Kurtz drafted in 1975 a statement attacking astrology; the statement was signed by 192 leading scientists, including 19 Nobel prize winners. The statement raises three main issues: astrology originated as part of a magical world view, the planets are too distant for there to be any physical foundation for astrology, and people believe it merely out of longing for comfort ([2], pp. 9f.). None of these objections is ground for condemning astrology as pseudoscience. To show this, I shall briefly discuss articles written by Bok [1] and Jerome [12] in support of the statement.

According to Bok, to work on statistical tests of astrological predictions is a waste of time unless it is demonstrated that astrology has some sort of physical foundation ([1], p. 31). He uses the smallness of gravitational and radiative effects of the stars and planets to suggest that there is no such foundation. He also discusses the psychology of belief in astrology, which is the result of individuals' desperation in seeking solutions to their serious personal problems. Jerome devotes most of his article to the origins of astrology in the magical principle of correspondences. He claims that astrology is a system of magic rather than science, and that it fails "not because of any inherent inaccuracies due to precession or lack of exact knowledge concerning time of birth or conception, but rather because its interpretations and predictions are grounded in the ancients' magical world view" ([12], p. 46). He does however discuss some statistical tests of astrology, which I shall return to below.

These objections do not show that astrology is a pseudoscience. First, origins are irrelevant to scientific status. The alchemical origins of chemistry ([11], pp. 10–18) and the occult beginnings of medicine [8] are as magical as those of astrology, and historians have detected mystical influences in the work of many great scientists, including Newton and Einstein. Hence astrology cannot be condemned simply for the magical origins of its principles. Similarly, the psychology of popular belief is also in itself irrelevant to the status of astrology: people often believe even good theories for illegitimate reasons, and even if most people believe astrology for personal, irrational reasons, good reasons may be available.¹ Finally the lack of a physical foundation hardly marks a theory as unscientific ([22], p. 2). Examples: when Wegener [31] proposed continental drift, no mechanism was known, and a link between smoking and cancer has been established statistically [28] though the details of carcinogenesis remain to be discovered. Hence the objections of Bok, Jerome and Kurtz fail to mark astrology as pseudoscience.

Now we must consider the application of the criteria of verifiability and falsifiability to astrology. Roughly, a theory is said to be verifiable if it is possible to deduce observation statements from it. Then in principle, observations can be used to confirm or disconfirm the theory. A theory is

scientific only if it is verifiable. The vicissitudes of the verification principle are too well known to recount here ([9], ch. 4). Attempts by A. J. Ayer to articulate the principle failed either by ruling out most of science as unscientific, or by ruling out nothing. Moreover, the theory/observation distinction has increasingly come into question. All that remains is a vague sense that testability somehow is a mark of scientific theories ([9], ch. 4; [10], pp. 30-32).

Well, astrology is vaguely testable. Because of the multitude of influences resting on tendencies rather than laws, astrology is incapable of making precise predictions. Nevertheless, attempts have been made to test the reality of these alleged tendencies, using large scale surveys and statistical evaluation. The pioneer in this area was Michel Gauquelin, who examined the careers and times of birth of 25,000 Frenchmen. Astrology suggests that people born under certain signs or planets are likely to adopt certain occupations: for example, the influence of the warlike planet Mars tends to produce soldiers or athletes, while Venus has an artistic influence. Notably, Gauquelin found *no significant correlation* between careers and either sun sign, moon sign, or ascendant sign. However, he did find some statistically interesting correlations between certain occupations of people and the position of certain planets at the time of their birth ([5], ch. 11, [6]). For example, just as astrology would suggest, there is a greater than chance association of athletes and Mars, and a greater than chance association of scientists and Saturn, where the planet is rising or at its zenith at the moment of the individual's birth.

These findings and their interpretation are highly controversial, as are subsequent studies in a similar vein [7]. Even if correct, they hardly verify astrology, especially considering the negative results found for the most important astrological categories. I have mentioned Gauquelin in order to suggest that through the use of statistical techniques astrology is at least *verifiable*. Hence the verification principle does not mark astrology as pseudoscience.

Because the predictions of astrologers are generally vague, a Popperian would assert that the real problem with astrology is that it is not falsifiable: astrologers cannot make predictions which if unfulfilled would lead them to give up their theory. Hence because it is unfalsifiable, astrology is unscientific.

But the doctrine of falsifiability faces serious problems as described by Duhem [4], Quine [21], and Lakatos [15]. Popper himself noticed early that no observation ever guarantees falsification: a theory can always be retained by introducing or modifying auxiliary hypotheses, and even observation statements are not incorrigible ([17], p. 50). Methodological decisions about what can be tampered with are required to block the escape from falsification. However, Lakatos has persuasively argued that making such decisions in advance of tests is arbitrary and may often lead to

overhasty rejection of a sound theory which *ought* to be saved by anti-falsificationist stratagems ([15], pp. 112 ff.). Falsification only occurs when a better theory comes along. Then falsifiability is only a matter of replaceability by another theory, and since astrology is in principle replaceable by another theory, falsifiability provides no criterion for rejecting astrology as pseudoscientific. We saw in the discussion of Gauquelin that astrology can be used to make predictions about statistical regularities, but the non-existence of these regularities does not falsify astrology; but here astrology does not appear worse than the best of scientific theories, which also resist falsification until alternative theories arise.²

Astrology cannot be condemned as pseudoscientific on the grounds proposed by verificationists, falsificationists, or Bok and Jerome. But undoubtedly astrology today faces a great many unsolved problems ([32], ch. 5). One is the negative result found by Gauquelin concerning careers and signs. Another is the problem of the precession of the equinoxes, which astrologers generally take into account when heralding the "Age of Aquarius" but totally neglect when figuring their charts. Astrologers do not always agree on the significance of the three planets, Neptune, Uranus and Pluto, that were discovered since Ptolemy. Studies of twins do not show similarities of personality and fate that astrology would suggest. Nor does astrology make sense of mass disasters, where numerous individuals with very different horoscopes come to similar ends.

But problems such as these do not in themselves show that astrology is either false or pseudoscientific. Even the best theories face unsolved problems throughout their history. To get a criterion demarcating astrology from science, we need to consider it in a wider historical and social context.

A demarcation criterion requires a matrix of three elements: theory, community, historical context. Under the first heading, "theory", fall familiar matters of structure, prediction, explanation and problem solving. We might also include the issue raised by Bok and Jerome about whether the theory has a physical foundation. Previous demarcationists have concentrated on this theoretical element, evident in the concern of the verification and falsification principles with prediction. But we have seen that this approach is not sufficient for characterizing astrology as pseudoscientific.

We must also consider the *community* of advocates of the theory, in this case the community of practitioners of astrology. Several questions are important here. First, are the practitioners in agreement on the principles of the theory and on how to go about solving problems which the theory faces? Second, do they care, that is, are they concerned about explaining anomalies and comparing the success of their theory to the record of other theories? Third, are the practitioners actively involved in attempts at confirming and disconfirming their theory?

The question about comparing the success of a theory with that of other theories introduces the third element of the matrix, historical context. The historical work of Kuhn and others has shown that in general a theory is rejected only when (1) it has faced anomalies over a long period of time and (2) it has been challenged by another theory. Hence under the heading of historical context we must consider two factors relevant to demarcation: the record of a theory over time in explaining new facts and dealing with anomalies, and the availability of alternative theories.

We can now propose the following principle of demarcation:

A theory or discipline which purports to be scientific is *pseudoscientific* if and only if:

- 1 it has been less progressive than alternative theories over a long period of time, and faces many unsolved problems; but
- 2 the community of practitioners makes little attempt to develop the theory towards solutions of the problems, shows no concern for attempts to evaluate the theory in relation to others, and is selective in considering confirmations and disconfirmations.

Progressiveness is a matter of the success of the theory in adding to its set of facts explained and problems solved ([15], p. 118; cf. [26], p. 83).

This principle captures, I believe, what is most importantly unscientific about astrology. First, astrology is dramatically unprogressive, in that it has changed little and has added nothing to its explanatory power since the time of Ptolemy. Second, problems such as the precession of equinoxes are outstanding. Third, there are alternative theories of personality and behavior available: one need not be an uncritical advocate of behaviorist, Freudian, or Gestalt theories to see that since the nineteenth century psychological theories have been expanding to deal with many of the phenomena which astrology explains in terms of heavenly influences. The important point is not that any of these psychological theories is established or true, only that they are growing alternatives to a long-static astrology. Fourth and finally, the community of astrologers is generally unconcerned with advancing astrology to deal with outstanding problems or with evaluating the theory in relation to others.³ For these reasons, my criterion marks astrology as pseudoscientific.*

This demarcation criterion differs from those implicit in Lakatos and

* Since writing this paper, Thagard has offered a revised account of pseudoscience in chapter 9 of his book *Computational Philosophy of Science* (Cambridge, Mass.: MIT Press, 1988). This revised account is discussed in our commentary on chapter 1.

Kuhn. Lakatos has said that what makes a series of theories constituting a research program scientific is that it is progressive: each theory in the series has greater corroborated content than its predecessor ([15], p. 118). While I agree with Lakatos that progressiveness is a central notion here, it is not sufficient to distinguish science from pseudoscience. We should not brand a nonprogressive discipline as pseudoscientific unless it is being maintained against more progressive alternatives. Kuhn's discussion of astrology focuses on a different aspect of my criterion. He says that what makes astrology unscientific is the absence of the paradigm-dominated puzzle solving activity characteristic of what he calls normal science ([14], p. 9). But as Watkins has suggested, astrologers are in some respects model normal scientists: they concern themselves with solving puzzles at the level of individual horoscopes, unconcerned with the foundations of their general theory or paradigm ([30], p. 32). Hence that feature of normal science does not distinguish science from pseudoscience. What makes astrology pseudoscientific is not that it lacks periods of Kuhnian normal science, but that its proponents adopt uncritical attitudes of "normal" scientists despite the existence of more progressive alternative theories. (Note that I am not agreeing with Popper [18] that Kuhn's normal scientists are unscientific; they can become unscientific only when an alternative paradigm has been developed.) However, if one looks not at the puzzle solving at the level of particular astrological predictions, but at the level of theoretical problems such as the precession of the equinoxes, there is some agreement between my criterion and Kuhn's; astrologers do not have a paradigm-induced confidence about solving theoretical problems.

Of course, the criterion is intended to have applications beyond astrology. I think that discussion would show that the criterion marks as pseudoscientific such practices as witchcraft and pyramidology, while leaving contemporary physics, chemistry and biology unthreatened. The current fad of biorhythms, implausibly based like astrology on date of birth, cannot be branded as pseudoscientific because we lack alternative theories giving more detailed accounts of cyclical variations in human beings, although much research is in progress.⁴

One interesting consequence of the above criterion is that a theory can be scientific at one time but pseudoscientific at another. In the time of Ptolemy or even Kepler, astrology had few alternatives in the explanation of human personality and behavior. Existing alternatives were scarcely more sophisticated or corroborated than astrology. Hence astrology should be judged as not pseudoscientific in classical or Renaissance times, even though it is pseudoscientific today. Astrology was not simply a perverse sideline of Ptolemy and Kepler, but part of their scientific activity, even if a physicist involved with astrology today should be looked at askance. Only when the historical and social aspects of science are neglected does it become plausible that pseudoscience is an unchanging category. Ration-

ality is not a property of ideas eternally: ideas, like actions, can be rational at one time but irrational at others. Hence relativizing the science/pseudoscience distinction to historical periods is a desirable result.

But there remains a challenging historical problem. According to my criterion, astrology only became pseudoscientific with the rise of modern psychology in the nineteenth century. But astrology was already virtually excised from scientific circles by the beginning of the eighteenth. How could this be? The simple answer is that a theory can take on the appearance of an unpromising project well before it deserves the label of pseudoscience. The Copernican revolution and the mechanism of Newton, Descartes and Hobbes undermined the plausibility of astrology.⁵ Lynn Thorndike [27] has described how the Newtonian theory pushed aside what had been accepted as a universal natural law, that inferiors such as inhabitants of earth are ruled and governed by superiors such as the stars and the planets. William Stahlman [23] has described how the immense growth of science in the seventeenth century contrasted with the stagnation of astrology. These developments provided good reasons for discarding astrology as a promising pursuit, but they were not yet enough to brand it as pseudoscientific, or even to refute it.

Because of its social aspect, my criterion might suggest a kind of cultural relativism. Suppose there is an isolated group of astrologers in the jungles of South America, practicing their art with no awareness of alternatives. Are we to say that astrology is *for them* scientific? Or, going in the other direction, should we count as alternative theories ones which are available to extraterrestrial beings, or which someday will be conceived? This wide construal of "alternative" would have the result that our best current theories are probably pseudoscientific. These two questions employ, respectively, a too narrow and a too broad view of alternatives. By an alternative theory I mean one generally available in the world. This assumes first that there is some kind of communication network to which a community has, or should have, access. Second, it assumes that the onus is on individuals and communities to find out about alternatives. I would argue (perhaps against Kuhn) that this second assumption is a general feature of rationality; it is at least sufficient to preclude ostrichism as a defense against being judged pseudoscientific.

In conclusion, I would like to say why I think the question of what constitutes a pseudoscience is important. Unlike the logical positivists, I am not grinding an anti-metaphysical ax, and unlike Popper, I am not grinding an anti-Freudian or anti-Marxian one.⁶ My concern is social: society faces the twin problems of lack of public concern with the advancement of science, and lack of public concern with the important ethical issues now arising in science and technology, for example around the topic of genetic engineering. One reason for this dual lack of concern is the wide popularity of pseudoscience and the occult among the general

public. Elucidation of how science differs from pseudoscience is the philosophical side of an attempt to overcome public neglect of genuine science.⁷

■ | Notes

1. However, astrology would doubtlessly have many fewer supporters if horoscopes tended less toward compliments and pleasant predictions and more toward the kind of analysis included in the following satirical horoscope from the December, 1977, issue of *Mother Jones*: VIRGO (Aug. 23–Sept. 22). You are the logical type and hate disorder. This nitpicking is sickening to your friends. You are cold and unemotional and sometimes fall asleep while making love. Virgos make good bus drivers.
2. For an account of the comparative evaluation of theories, see [26].
3. There appear to be a few exceptions; see [32].
4. The fad of biorhythms, now assuming a place beside astrology in the popular press, must be distinguished from the very interesting work of Frank Brown and others on biological rhythms. For a survey, see [5].
5. Plausibility is in part a matter of a hypothesis being of an appropriate *kind*, and is relevant even to the acceptance of a theory. See [26], p. 90, and [25].
6. On psychoanalysis see [3]. I would argue that Cioffi neglects the question of alternatives to psychoanalysis and the question of its progressiveness.
7. I am grateful to Dan Hausman and Elias Baumgarten for comments.

■ | References

- [1] Bok, Bart J. "A Critical Look at Astrology." In [2]. Pages 21–33.
- [2] ———, Jerome, Lawrence E., and Kurtz, Paul. *Objections to Astrology*. Buffalo: Prometheus Books, 1975.
- [3] Cioffi, Frank. "Freud and the Idea of a Pseudoscience." In *Explanation in the Behavioral Sciences*. Edited by R. Borger and F. Cioffi. Cambridge: Cambridge University Press, 1970. Pages 471–499.
- [4] Duhem, P. *The Aim and Structure of Physical Theory*. (trans.) P. Wiener. New York: Atheneum, 1954. (Translated from 2nd edition of *La Théorie physique: son objet et sa structure*. Paris: Marcel Rivière & Cie, 1914.)
- [5] Gauquelin, Michel. *The Cosmic Clocks*. Chicago: Henry Regnery, 1967.
- [6] ———. *The Scientific Basis of Astrology*. New York: Stein and Day, 1969.
- [7] ———. "The Zelen Test of the Mars Effect." *The Humanist* 37 (1977): 30–35.

- [8] Haggard, Howard W. *Mystery, Magic, and Medicine*. Garden City: Doubleday, Doran & Company, 1933.
- [9] Hempel, Carl. *Aspects of Scientific Explanation*. New York: The Free Press, 1965.
- [10] ———. *Philosophy of Natural Science*. Englewood Cliffs: Prentice-Hall, 1966.
- [11] Ihde, Aaron J. *The Development of Modern Chemistry*. New York: Harper and Row, 1964.
- [12] Jerome, Lawrence E. "Astrology: Magic or Science?" In [2]. Pages 37–62.
- [13] Koestler, Arthur. *The Sleepwalkers*. Harmondsworth: Penguin, 1964.
- [14] Kuhn, T. S. "Logic of Discovery or Psychology of Research?" In [16]. Pages 1–23.
- [15] Lakatos, Imre. "Falsification and the Methodology of Scientific Research Programmes." In [16]. Pages 91–195.
- [16] ——— and Musgrave, Alan. (eds.). *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press, 1970.
- [17] Popper, Karl. *The Logic of Scientific Discovery*. London: Hutchinson, 1959. (Originally published as *Logik der Forschung*. Vienna: J. Springer, 1935.)
- [18] ———. "Normal Science and its Dangers." In [16]. Pages 51–58.
- [19] Ptolemy. *The Almagest (The Mathematical Composition)*. (As printed in Hutchins, Robert Maynard (ed.). *Great Books of the Western World*, Volume 16. Chicago: Encyclopedia Britannica, Inc., 1952. Pages 1–478.)
- [20] ———. *Tetrabiblos*. Edited and translated by F. E. Robbins. Cambridge: Harvard University Press, 1940.
- [21] Quine, W. V. O. "Two Dogmas of Empiricism." In *From a Logical Point of View*. New York: Harper & Row, 1963. Pages 20–46. (Originally published in *The Philosophical Review* 60 (1951): 20–43.)
- [22] Sagan, Carl. "Letter." *The Humanist* 36 (1976): 2.
- [23] Stahlman, William D. "Astrology in Colonial America: An Extended Query." *William and Mary Quarterly* 13 (1956): 551–563.
- [24] Swift, Jonathan. "The Partridge Papers." In *The Prose Works of Jonathan Swift*, Volume 2. Oxford: Basil Blackwell, 1940–1968. Pages 139–170.
- [25] Thagard, Paul R. "The Autonomy of a Logic of Discovery." In L. W. Sumner et al., eds., *Pragmatism and Purpose: Essays Presented to Thomas A. Goudge* (Toronto: University of Toronto Press, 1981), 248–60.
- [26] ———. "The Best Explanation: Criteria for Theory Choice." *Journal of Philosophy* 75 (1978): 76–92.
- [27] Thorndike, Lynn. "The True Place of Astrology in the History of Science." *Isis* 46 (1955): 273–278.
- [28] U.S. Department of Health, Education and Welfare. *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service*. Washington, D.C.: U.S. Government Printing Office, 1964.

- [29] Voltaire. "Astrologie" and "Astronomie." *Dictionnaire Philosophique*. In *Oeuvres Complètes de Voltaire*, Volume XVII. Paris: Garnier Frères, 1878–1885. Pages 446–453.
- [30] Watkins, J. W. N. "Against 'Normal Science'." In [16]. Pages 25–37.
- [31] Wegener, Alfred. "Die Entstehung der Kontinente." *Petermanns Geographische Mittheilung* 58 (1912): 185–195, 253–256, 305–309.
- [32] West, J. A. and Toonder, J. G. *The Case for Astrology*. Harmondsworth: Penguin, 1973.

Creation-Science Is Not Science

In December 1981 I appeared as an expert witness for the plaintiffs and the American Civil Liberties Union (ACLU) in their successful challenge of Arkansas Act 590, which demanded that teachers give "balanced treatment" to "creation-science" and evolutionary ideas.¹ My presence occasioned some surprise, for I am an historian and philosopher of science. In this essay, I do not intend to apologize for either my existence or my calling, nor do I intend to relive past victories²; rather, I want to explain why a philosopher and historian of science finds the teaching of "creation-science" in science classrooms offensive.

Obviously, the crux of the issue—the center of the plaintiffs' case—is the status of creation-science. Its advocates claim that it is genuine science and may, therefore, be legitimately and properly taught in the public schools. Its detractors claim that it is not genuine science but a form of religion—dogmatic Biblical literalism by another name. Which is it, and who is to decide?

It is somewhat easier to describe who should participate in decisions on this issue. On the one hand, one naturally appeals to the authority of religious people and theologians. Does creation-science fit the accepted definitions of a religion? (In Arkansas, the ACLU produced theologians who said that indeed it did.) One also appeals to the authority of scientists. Does creation-science fit current definitions of science? (In Arkansas, the ACLU produced scientists who said that indeed it did not.)³

Having, as it were, appealed to the practitioners—theologians and scientists—a link still seems to be missing. Someone is needed to talk at a more theoretical level about the nature of science—any science—and then show that creation-science simply does not fit the part. As a philosopher and an historian, it is my job to look at science, and to ask precisely those questions about defining characteristics.

■ | What Is Science?

It is simply not possible to give a neat definition—specifying necessary and sufficient characteristics—which separates all and only those things that have ever been called “science.” The concept “science” is not as easily definable as, for example, the concept “triangle.” Science is a phenomenon that has developed through the ages—dragging itself apart from religion, philosophy, superstition, and other bodies of human opinion and belief.⁴

What we call “science” today is a reasonably striking and distinctive set of claims, which have a number of characteristic features. As with most things in life, some items fall on the borderline between science and nonscience (e.g., perhaps Freudian psychoanalytic theory). But it is possible to state positively that, for example, physics and chemistry are sciences, and Plato’s Theory of Forms and Swedenborgian theology are not.⁵

In looking for defining features, the obvious place to start is with science’s most striking aspect—it is an empirical enterprise about the real world of sensation. This is not to say that science refers only to observable entities. Every mature science contains unobservables, like electrons and genes, but ultimately, they refer to the world around us. Science attempts to understand this empirical world. What is the basis for this understanding? Surveying science and the history of science today, one thing stands out: science involves a search for order. More specifically, science looks for unbroken, blind, natural regularities (*laws*). Things in the world do not happen in just any old way. They follow set paths, and science tries to capture this fact. Bodies of science, therefore, known variously as “theories” or “paradigms” or “sets of models,” are collections of laws.⁶

Thus, in Newtonian physics we find Newton’s three laws of motion, the law of gravitational attraction, Kepler’s laws of planetary motion, and so forth. Similarly, for instance, in population genetics we find the Hardy-Weinberg law. However, when we turn to something like philosophy, we do not find the same appeal to empirical law. Plato’s Theory of Forms only indirectly refers to this world. Analogously, religion does not insist on unbroken law. Indeed, religious beliefs frequently allow or suppose events outside law or else events that violate law (miracles). Jesus feeding the 5,000 with the loaves and fishes was one such event. This is not to say that religion is false, but it does say that religion is not science. When the loaves and fishes multiplied to a sufficiency to feed so many people, things happened that did not obey natural law, and hence the feeding of the 5,000 is an event beyond the ken of science.⁷

A major part of the scientific enterprise involves the use of law to effect *explanation*. One tries to show why things are as they are—and how they fall beneath or follow from law (together perhaps with certain specified initial conditions). Why, for example, does a cannon ball go in a

parabola and not in a circle? Because of the constraints of Newton's laws. Why do two blue-eyed parents always have blue-eyed children? Because this trait obeys Mendel's first law, given the particular way in which the genes control eye-color. A scientific explanation must appeal to law and must show that what is being explained had to occur. The explanation excludes those things that did not happen.⁸

The other side of explanation is *prediction*. The laws indicate what is going to happen: that the ball will go in a parabola, that the child will be blue-eyed. In science, as well as in futurology, one can also, as it were, predict backwards. Using laws, one infers that a particular, hitherto-unknown phenomenon or event took place in the past. Thus, for instance, one might use the laws of physics to infer back to some eclipse of the sun reported in ancient writings.

Closely connected with the twin notions of explanation and prediction comes *testability*. A genuine scientific theory lays itself open to check against the real world: the scientist can see if the inferences made in explanation and prediction actually obtain in nature. Does the chemical reaction proceed as suspected? In Young's double slit experiment, does one find the bands of light and dark predicted by the wave theory? Do the continents show the expected after-effects of drift?

Testability is a two-way process. The researcher looks for some positive evidence, for *confirmation*. No one will take seriously a scientific theory that has no empirical support (although obviously a younger theory is liable to be less well-supported than an older theory). Conversely, a theory must be open to possible refutation. If the facts speak against a theory, then it must go. A body of science must be *falsifiable*. For example, Kepler's laws could have been false: if a planet were discovered going in squares, then the laws would have been shown to be incorrect. However, in distinguishing science from nonscience, no amount of empirical evidence can disprove, for example, the Kantian philosophical claim that one *ought* to treat people as ends rather than means. Similarly, Catholic religious claims about transubstantiation (the changing of the bread and wine into the body and blood of Christ) are unfalsifiable.⁹

Science is *tentative*. Ultimately, a scientist must be prepared to reject his theory. Unfortunately, not all scientists are prepared to do in practice what they promise to do in theory; but the weaknesses of individuals are counterbalanced by the fact that, as a group, scientists do give up theories that fail to answer to new or reconsidered evidence. In the last 30 years, for example, geologists have reversed their strong convictions that the continents never move.

Scientists do not, of course, immediately throw their theories away as soon as any counter-evidence arrives. If a theory is powerful and successful, then some problems will be tolerated, but scientists must be prepared to change their minds in the face of the empirical evidence. In this regard, the scientists differ from both the philosophers and the theologians. Noth-

ing in the real world would make the Kantian change his mind, and the Catholic is equally dogmatic, despite any empirical evidence about the stability of bread and wine. Such evidence is simply considered irrelevant.¹⁰

Some other features of science should also be mentioned, for instance, the urge for simplicity and unification; however, I have now listed the major characteristics. Good science—like good philosophy and good religion—presupposes an attitude that one might describe as professional *integrity*. A scientist should not cheat or falsify data or quote out of context or do any other thing that is intellectually dishonest. Of course, as always, some individuals fail; but science as a whole disapproves of such actions. Indeed, when transgressors are detected, they are usually expelled from the community. Science depends on honesty in the realm of ideas. One may cheat on one's taxes; one may not fiddle the data.¹¹

■ | Creation-Science Considered

How does creation-science fit the criteria of science listed in the previous section? By "creation-science" in this context, I refer not just to the definition given in Act 590, but to the whole body of literature which goes by that name. The doctrine includes the claims that the universe is very young (6,000 to 20,000 years), that everything started instantaneously, that human beings had ancestry separate from apes, and that a monstrous flood once engulfed the entire earth.¹²

LAWS—NATURAL REGULARITIES

Science is about unbroken, natural regularity. It does not admit miracles. It is clear, therefore, that again and again, creation-science invokes happenings and causes outside of law. For instance, the only reasonable inference from Act 590 (certainly the inference that was accepted in the Arkansas court) is that for creation-science the origin of the universe and life in it is not bound by law. Whereas the definition of creation-science includes the unqualified phrase "sudden creation of the universe, energy and life from nothing," the definition of evolution specifically includes the qualification that its view of origins is "naturalistic." Because "naturalistic" means "subject to empirical law," the deliberate omission of such a term in the characterization of creation-science means that no laws were involved.

In confirmation of this inference, we can find identical claims in the writings of creation scientists: for instance, the following passage from Duane T. Gish's popular work *Evolution—The Fossils Say No!*

CREATION. By creation we mean the bringing into being of the basic kinds of plants and animals by the process of sudden, or fiat, creation described in the first two chapters of Genesis. Here we find the creation by God of the plants and animals, each commanded to reproduce after its own kind using processes which were essentially instantaneous.

We do not know how God created, what processes He used, *for God used processes which are not now operating anywhere in the natural universe*. This is why we refer to divine creation as special creation. We cannot discover by scientific investigations anything about the creative processes used by God.¹³

By Gish's own admission, we are not dealing with science. Similar sentiments can be found in *The Genesis Flood* by John Whitcomb, Jr., and Henry M. Morris:

But during the period of Creation, God was introducing order and organization and energization into the universe in a very high degree, even to life itself. *It is thus quite plain that the processes used by God in creation were utterly different from the processes which now operate in the universe!* The Creation was a unique period, entirely incommensurate with this present world. This is plainly emphasized and reemphasized in the divine revelation which God has given us concerning Creation, which concludes with these words: 'And the heavens and the earth were *finished*, and *all* the host of them. And on the seventh day God *finished* His work which He had made; and He *rested* on the seventh day from *all* His work which He had made. And God blessed the seventh day, and hallowed it; because that in it He *rested* from *all* his work which God had created and made.' In view of these strong and repeated assertions, is it not the height of presumption for man to attempt to study Creation in terms of present processes?¹⁴

Creation scientists generally acknowledge this work to be *the* seminal contribution that led to the growth of the creation-science movement. Morris, in particular, is the father figure of creation-science and Gish his chief lieutenant.

Creation scientists also break with law in many other instances. The creationists believe that the Flood, for example, could not have just occurred through blind regularities. As Whitcomb and Morris make very clear, certain supernatural interventions were necessary to bring about the Flood.¹⁵ Similarly, in order to ensure the survival of at least some organisms, God had to busy himself and break through law.

EXPLANATION AND PREDICTION

Given the crucial role that law plays for the scientist in these processes, neither explanation nor prediction is possible where no law exists.

Thus, explanation and prediction simply cannot even be attempted when one deals with creation-science accounts either of origins or of the Flood.

Even against the broader vistas of biology, creation-science is inadequate. Scientific explanation/prediction must lead to the thing being explained/predicted, showing why that thing obtains and not other things. Why does the ball go in a parabola? Why does it not describe a circle? Take an important and pervasive biological phenomenon, namely, "homologies," the isomorphisms between the bones of different animals. These similarities were recognized as pervasive facets of nature even before Darwin published *On the Origin of Species*. Why are the bones in the forelimbs of men, horses, whales, and birds all so similar, even though the functions are quite different? Evolutionists explain homologies naturally and easily, as a result of common descent. Creationists can give no explanation, and make no predictions. All they can offer is the disingenuous comment that homology signifies nothing, because classification is all man-made and arbitrary anyway. Is it arbitrary that man is not classified with the birds?¹⁶ Why are Darwin's finches distributed in the way that we find on the Galapagos? Why are there 14 separate species of this little bird, scattered over a small group of islands in the Pacific on the equator? On those rare occasions when Darwin's finches do fly into the pages of creation-science, it is claimed either that they are all the same species (false), or that they are a case of degeneration from one "kind" created back at the beginning of life.¹⁷ Apart from the fact that "kind" is a term of classification to be found only in Genesis, this is no explanation. How could such a division of the finches have occurred, given the short span that the creationists allow since the Creation? And, in any case, Darwin's finches are anything but degenerates. Different species of finch have entirely different sorts of beaks, adapted for different foodstuffs—evolution of the most sophisticated type.¹⁸

TESTABILITY, CONFIRMATION, AND FALSIFIABILITY

Testability, confirmation, and falsifiability are no better treated by creation-science. A scientific theory must provide more than just after-the-fact explanations of things that one already knows. One must push out into the frontiers of new knowledge, trying to predict new facts, and risking the theory against the discovery of possible falsifying information. One cannot simply work at a secondary level, constantly protecting one's views against threat: forever inventing *ad hoc* hypotheses to save one's core assumptions.

Creation scientists do little or nothing by way of genuine test. Indeed, the most striking thing about the whole body of creation-science literature is the virtual absence of any experimental or observational work by creation scientists. Almost invariably, the creationists work exclusively with the discoveries and claims of evolutionists, twisting the conclusions to their

own ends. Argument proceeds by showing evolution (specifically Darwinism) wrong, rather than by showing Creationism right.

However, this way of proceeding—what the creationists refer to as the “two model approach”—is simply a fallacious form of argument. The views of people like Fred Hoyle and N. C. Wickramasinghe, who believe that life comes from outer space, are neither creationist nor truly evolutionist.¹⁹ Denying evolution in no way proves Creationism. And, even if a more straightforward either/or between evolution and Creationism existed, the perpetually negative approach is just not the way that science proceeds. One must find one’s own evidence in favor of one’s position, just as physicists, chemists, and biologists do.

Do creation scientists ever actually expose their theories and ideas to test? Even if they do, when new counter-empirical evidence is discovered, creation scientists appear to pull back, refusing to allow their position to be falsified.

Consider, for instance, the classic case of the “missing link”—namely, that between man and his ancestors. The creationists say that there are no plausible bridging organisms whatsoever. Thus, this super-gap between man and all other animals (alive or dead) supposedly underlines the creationists’ contention that man and apes have separate ancestry. But what about the australopithecines, organisms that paleontologists have, for most of this century, claimed are plausible human ancestors? With respect, argue the creationists, australopithecines are **not** links, because they had ape-like brains, they *walked* like apes, and they used their knuckles for support, just like gorillas. Hence, the gap remains.²⁰

However, such a conclusion can be maintained only by blatant disregard of the empirical evidence. *Australopithecus afarensis* was a creature with a brain the size of that of an ape which walked upright.²¹ Yet the creationists do not concede defeat. They then argue that the *Australopithecus afarensis* is like an orangutan.²² In short, nothing apparently makes the creationists change their minds, or allows their views to be tested, lest they be falsified.

TENTATIVENESS

Creation-science is not science because there is absolutely no way in which creationists will budge from their position. Indeed, the leading organization of creation-science, The Creation Research Society (with 500 full members, all of whom must have an advanced degree in a scientific/technological area), demands that its members sign a statement affirming that they take the Bible as literally true.²³ Unfortunately, an organization cannot require such a condition of membership, and then claim to be a scientific organization. Science must be open to change, however confident one may feel at present. Fanatical dogmatism is just not acceptable.

INTEGRITY

Creation scientists use any fallacy in the logic books to achieve their ends. Most particularly, apart from grossly distorting evolutionists' positions, the creation-scientists frequently use inappropriate or incomplete quotations. They take the words of some eminent evolutionist, and attempt to make him or her say exactly the opposite to that intended. For instance, in *Creation: The Facts of Life*, author Gary E. Parker constantly refers to "noted Harvard geneticist" Richard Lewontin as claiming that the hand and the eye are the best evidence of God's design.²⁴ Can this reference really be true? Has the author of *The Genetic Basis of Evolutionary Change*²⁵ really fore sworn Darwin for Moses? In fact, when one looks at Lewontin's writings, one finds that he says that *before Darwin*, people believed the hand and the eye to be the effect of direct design. Today, scientists believe that such features were produced by the natural process of evolution through natural selection; but, a reader learns nothing of this from Parker's book.

What are the essential features of science? Does creation-science have any, all, or none of these features? My answer to this is none. By every mark of what constitutes science, creation-science fails. And, although it has not been my direct purpose to show its true nature, it is surely there for all to see. Miracles brought about by an intervening supervising force speak of only one thing. Creation "science" is actually dogmatic religious Fundamentalism. To regard it as otherwise is an insult to the scientist, as well as to the believer who sees creation-science as a blasphemous distortion of God-given reason. I believe that creation-science should not be taught in the public schools because creation-science is not science.

■ | Notes

1. In fact, Act 590 demanded that *if* one teach[es] evolution, *then* one must also teach creation-science. Presumably a teacher could have stayed away from origins entirely—albeit with large gaps in some courses.
2. For a brief personal account of my experiences, see Michael Ruse, "A Philosopher at the Monkey Trial," *New Scientist* (1982): 317–319.
3. Judge William Overton's ruling on the constitutionality (or, rather, unconstitutionality) of Act 590 gives a fair and full account of the various claims made by theologians (including historians and sociologists of religion) and scientists.
4. In my book, *The Darwinian Revolution: Science Red in Tooth and Claw* (Chicago, IL: University of Chicago Press, 1979), I look at the way science was breaking apart from religion in the 19th century.
5. What follows is drawn from a number of basic books in the philosophy of science, including R. B. Braithwaite, *Scientific Explanation* (Cambridge, England:

Cambridge University Press, 1953); Karl R. Popper, *The Logic of Scientific Discovery* (London: Hutchinson, 1959); E. Nagel, *The Structure of Science* (London: Routledge and Kegan Paul, 1961); Thomas S. Kuhn, *The Structure of Scientific Revolutions* (Chicago, IL: University of Chicago Press, 1962); and C. G. Hempel, *Philosophy of Natural Science* (Englewood Cliffs, NJ: Prentice-Hall, 1966). The discussion is the same as what I provided for the plaintiffs in a number of position papers. It also formed the basis of my testimony in court, and, as can be seen from Judge Overton's ruling, was accepted by the court virtually verbatim.

6. One sometimes sees a distinction drawn between "theory" and "model." At the level of this discussion, it is not necessary to discuss specific details. I consider various uses of these terms in my book, *Darwinism Defended: A Guide to the Evolution Controversies* (Reading, MA: Addison-Wesley, 1982).

7. For more on science and miracles, especially with respect to evolutionary questions, see my *Darwinian Revolution*, *op. cit.*

8. The exact relationship between laws and what they explain has been a matter of much debate. Today, I think most would agree that the connection must be fairly tight—the thing being explained should follow. For more on explanation in biology see Michael Ruse, *The Philosophy of Biology* (London: Hutchinson, 1973); and David L. Hull, *Philosophy of Biological Science* (Englewood Cliffs, NJ: Prentice-Hall, 1974). A popular thesis is that explanation of laws involves deduction from other laws. A theory is a body of laws bound in this way: a so-called "hypothetico-deductive" system.

9. Falsifiability today has a high profile in the philosophical and scientific literature. Many scientists, especially, agree with Karl Popper, who has argued that falsifiability is the criterion demarcating science from non-science (see especially his *Logic of Scientific Discovery*). My position is that falsifiability is an important part, but only one part, of a spectrum of features required to demarcate science from non-science. For more on this point, see my *Is Science Sexist? And Other Problems in the Biomedical Sciences* (Dordrecht, Holland: D. Reidel Publishing Company, 1981).

10. At the Arkansas trial, in talking of the tentativeness of science, I drew an analogy in testimony between science and the law. In a criminal trial, one tries to establish guilt "beyond a reasonable doubt." If this can be done, then the criminal is convicted. But, if new evidence is ever discovered that might prove the convicted person innocent, cases can always be reopened. In science, too, scientists make decisions less formally but just as strongly—and get on with business, but cases (theories) can be reopened.

11. Of course, the scientist as citizen may run into problems here!

12. The key definitions in Arkansas Act 590, requiring "balanced treatment" in the public schools, are found in Section 4 [of the Act]. Section 4(a) does not specify exactly how old the earth is supposed to be, but in court a span of 6,000 to 20,000 years emerged in testimony.

The fullest account of the creation-science position is given in Henry M. Morris, ed., *Scientific Creationism* (San Diego, CA: Creation-Life Publishers, 1974).

13. Duane T. Gish, *Evolution—The Fossils Say No!* (San Diego, CA: Creation-Life Publishers, 1973), pp. 22–25, his italics.

14. John Whitcomb, Jr., and Henry M. Morris, *The Genesis Flood* (Philadelphia, PA: Presbyterian and Reformed Publishing Company, 1961), pp. 223–224, their italics.
15. *Ibid.*, p. 76.
16. See Morris, *op. cit.*, pp. 71–72, and my discussion in *Darwinism Defended*, *op. cit.*
17. For instance, in John N. Moore and H. S. Slusher, *Biology: A Search for Order in Complexity* (Grand Rapids, MI: Zondervan, 1977).
18. D. Lack, *Darwin's Finches* (Cambridge, England: Cambridge University Press, 1947).
19. Fred Hoyle and N. C. Wickramasinghe, *Evolution from Space* (London: Dent, 1981).
20. Morris, *op. cit.*, p. 173.
21. Donald Johanson and M. Edey, *Lucy: The Beginnings of Humankind* (New York, NY: Simon and Schuster, 1981).
22. Gary E. Parker, *Creation: The Facts of Life* (San Diego, CA: Creation-Life Publishers, 1979), p. 113.
23. For details of these statements, see [footnote] 7 in Judge Overton's ruling.
24. Parker, *op. cit.* See, for instance, pp. 55 and 144. The latter passage is worth quoting in full:

Then there's 'the marvelous fit of organisms to the environment,' the special adaptations of cleaner fish, woodpeckers, bombardier beetles, etc., etc.,—what Darwin called 'Difficulties with the Theory,' and what Harvard's Lewontin (1978) called 'the chief evidence of a Supreme Designer.' Because of their 'perfection of structure,' he says, organisms 'appear to have been carefully and artfully designed.'

The pertinent article by Richard Lewontin is "Adaptation," *Scientific American* (September 1978).

25. Richard C. Lewontin, *The Genetic Basis of Evolutionary Change* (New York, NY: Columbia University Press, 1974).

Commentary: Science at the Bar—Causes for Concern

In the wake of the decision in the Arkansas Creationism trial (*McLean v. Arkansas*)¹ the friends of science are apt to be relishing the outcome. The creationists quite clearly made a botch of their case and there can be little doubt that the Arkansas decision may, at least for a time, blunt legislative pressure to enact similar laws in other states. Once the dust has settled, however, the trial in general and Judge William R. Overton's ruling in particular may come back to haunt us; for, although the verdict itself is probably to be commended, it was reached for all the wrong reasons and by a chain of argument which is hopelessly suspect. Indeed, the ruling rests on a host of misrepresentations of what science is and how it works.

The heart of Judge Overton's Opinion is a formulation of "the essential characteristics of science." These characteristics serve as touchstones for contrasting evolutionary theory with Creationism; they lead Judge Overton ultimately to the claim, specious in its own right, that since Creationism is not "science," it must be religion. The Opinion offers five essential properties that demarcate scientific knowledge from other things: "(1) It is guided by natural law; (2) it has to be explanatory by reference to natural law; (3) it is testable against the empirical world; (4) its conclusions are tentative, i.e., are not necessarily the final word; and (5) it is falsifiable."

These fall naturally into two families: properties (1) and (2) have to do with lawlikeness and explanatory ability; the other three properties have to do with the fallibility and testability of scientific claims. I shall deal with the second set of issues first, because it is there that the most egregious errors of fact and judgment are to be found.

At various key points in the Opinion, Creationism is charged with being untestable, dogmatic (and thus non-tentative), and unfalsifiable. All

three charges are of dubious merit. For instance, to make the inter-linked claims that Creationism is neither falsifiable nor testable is to assert that Creationism makes no empirical assertions whatever. That is surely false. Creationists make a wide range of testable assertions about empirical matters of fact. Thus, as Judge Overton himself grants (apparently without seeing its implications), the creationists say that the earth is of very recent origin (say 6,000 to 20,000 years old); they argue that most of the geological features of the earth's surface are diluvial in character (i.e., products of the postulated worldwide Noachian deluge); they are committed to a large number of factual historical claims with which the Old Testament is replete; they assert the limited variability of species. They are committed to the view that, since animals and man were created at the same time, the human fossil record must be paleontologically co-extensive with the record of lower animals. It is fair to say that no one has shown how to reconcile such claims with the available evidence—evidence which speaks persuasively to a long earth history, among other things.

In brief, these claims are testable, they have been tested, and they have failed those tests. Unfortunately, the logic of the Opinion's analysis precludes saying any of the above. By arguing that the tenets of Creationism are neither testable nor falsifiable, Judge Overton (like those scientists who similarly charge Creationism with being untestable) deprives science of its strongest argument against Creationism. Indeed, if any doctrine in the history of science has ever been falsified, it is the set of claims associated with "creation-science." Asserting that Creationism makes no empirical claims plays directly, if inadvertently, into the hands of the creationists by immunizing their ideology from empirical confrontation. The correct way to combat Creationism is to confute the empirical claims it does make, not to pretend that it makes no such claims at all.

It is true, of course, that some tenets of Creationism are not testable in isolation (e.g., the claim that man emerged by a direct supernatural act of creation). But that scarcely makes Creationism "unscientific." It is now widely acknowledged that many scientific claims are not testable in isolation, but only when embedded in a larger system of statements, some of whose consequences can be submitted to test.

Judge Overton's third worry about Creationism centers on the issue of revisability. Over and over again, he finds Creationism and its advocates "unscientific" because they have "refuse[d] to change it regardless of the evidence developed during the course of the[ir] investigation." In point of fact, the charge is mistaken. If the claims of modern-day creationists are compared with those of their nineteenth-century counterparts, significant shifts in orientation and assertion are evident. One of the most visible opponents of Creationism, Stephen Gould, concedes that creationists have modified their views about the amount of variability allowed at the level of species change. Creationists do, in short, change their minds from time

to time. Doubtless they would credit these shifts to their efforts to adjust their views to newly emerging evidence, in what they imagine to be a scientifically respectable way.

Perhaps what Judge Overton had in mind was the fact that some of Creationism's core assumptions (e.g., that there was a Noachian flood, that man did not evolve from lower animals, or that God created the world) seem closed off from any serious modification. But historical and sociological researches on science strongly suggest that the scientists of any epoch likewise regard some of their beliefs as so fundamental as not to be open to repudiation or negotiation. Would Newton, for instance, have been tentative about the claim that there were forces in the world? Are quantum mechanicians willing to contemplate giving up the uncertainty relation? Are physicists willing to specify circumstances under which they would give up energy conservation? Numerous historians and philosophers of science (e.g., Kuhn, Mitroff, Feyerabend, Lakatos) have documented the existence of a certain degree of dogmatism about core commitments in scientific research and have argued that such dogmatism plays a constructive role in promoting the aims of science. I am not denying that there may be subtle but important differences between the dogmatism of scientists and that exhibited by many creationists; but one does not even begin to get at those differences by pretending that science is characterized by an uncompromising open-mindedness.

Even worse, the *ad hominem* charge of dogmatism against Creationism egregiously confuses doctrines with the proponents of those doctrines. Since no law mandates that creationists should be invited into the classroom, it is quite irrelevant whether they themselves are close-minded. The Arkansas statute proposed that Creationism be taught, not that creationists should teach it. What counts is the epistemic status of Creationism, not the cognitive idiosyncrasies of the creationists. Because many of the theses of Creationism are testable, the mind set of creationists has no bearing in law or in fact on the merits of Creationism.

What about the other pair of essential characteristics which the *McLean Opinion* cites, namely, that science is a matter of natural law and explainable by natural law? I find the formulation in the *Opinion* to be rather fuzzy; but the general idea appears to be that it is inappropriate and unscientific to postulate the existence of any process or fact which cannot be explained in terms of some known scientific laws—for instance, the creationists' assertion that there are outer limits to the change of species "cannot be explained by natural law." Earlier in the *Opinion*, Judge Overton also writes "there is no scientific explanation for these limits which is guided by natural law," and thus concludes that such limits are unscientific. Still later, remarking on the hypothesis of the Noachian flood, he says: "A worldwide flood as an explanation of the world's geology is not the product of natural law, nor can its occurrence be explained by natural law." Quite how Judge Overton knows that a worldwide flood "cannot"

be explained by the laws of science is left opaque; and even if we did not know how to reduce a universal flood to the familiar laws of physics, this requirement is an altogether inappropriate standard for ascertaining whether a claim is scientific. For centuries scientists have recognized a difference between establishing the existence of a phenomenon and explaining that phenomenon in a lawlike way. Our ultimate goal, no doubt, is to do both. But to suggest, as the *McLean* Opinion does repeatedly, that an existence claim (e.g., there was a worldwide flood) is unscientific until we have found the laws on which the alleged phenomenon depends is simply outrageous. Galileo and Newton took themselves to have established the existence of gravitational phenomena, long before anyone was able to give a causal or explanatory account of gravitation. Darwin took himself to have established the existence of natural selection almost a half-century before geneticists were able to lay out the laws of heredity on which natural selection depended. If we took the *McLean* Opinion criterion seriously, we should have to say that Newton and Darwin were unscientific; and, to take an example from our own time, it would follow that plate tectonics is unscientific because we have not yet identified the laws of physics and chemistry which account for the dynamics of crustal motion.

The real objection to such creationist claims as that of the (relative) invariability of species is not that such invariability has not been explained by scientific laws, but rather that the evidence for invariability is less robust than the evidence for its contrary, variability. But to say as much requires renunciation of the Opinion's other charge—to wit, that Creationism is not testable.

I could continue with this tale of woeful fallacies in the Arkansas ruling, but that is hardly necessary. What is worrisome is that the Opinion's line of reasoning—which neatly coincides with the predominant tactic among scientists who have entered the public fray on this issue—leaves many loopholes for the creationists to exploit. As numerous authors have shown, the requirements of testability, revisability, and falsifiability are exceedingly *weak* requirements. Leaving aside the fact that (as I pointed out above) it can be argued that Creationism already satisfies these requirements, it would be easy for a creationist to say the following: "I will abandon my views if we find a living specimen of a species intermediate between man and apes." It is, of course, extremely unlikely that such an individual will be discovered. But, in that statement the creationist would satisfy, in one fell swoop, all the formal requirements of testability, falsifiability, and revisability. If we set very weak standards for scientific status—and, let there be no mistake, I believe that all of the Opinion's last three criteria fall in this category—then it will be quite simple for Creationism to qualify as "scientific."

Rather than taking on the creationists obliquely and in wholesale fashion by suggesting that what they are doing is "unscientific" *tout court*

(which is doubly silly because few authors can even agree on what makes an activity scientific), we should confront their claims directly and in piecemeal fashion by asking what evidence and arguments can be marshalled for and against each of them. The core issue is not whether Creationism satisfies some undemanding and highly controversial definitions of what is scientific; the real question is whether the existing evidence provides stronger arguments for evolutionary theory than for Creationism. Once that question is settled, we will know what belongs in the classroom and what does not. Debating the scientific status of Creationism (especially when "science" is construed in such an unfortunate manner) is a red herring that diverts attention away from the issues that should concern us.

Some defenders of the scientific orthodoxy will probably say that my reservations are just nitpicking ones, and that—at least to a first order of approximation—Judge Overton has correctly identified what is fishy about Creationism. The apologists for science, such as the editor of *The Skeptical Inquirer*, have already objected to those who criticize this whitewash of science "on arcane, semantic grounds . . . [drawn] from the most remote reaches of the academic philosophy of science."² But let us be clear about what is at stake. In setting out in the *McLean* Opinion to characterize the "essential" nature of science, Judge Overton was explicitly venturing into philosophical terrain. His *obiter dicta* are about as remote from well-founded opinion in the philosophy of science as Creationism is from respectable geology. It simply will not do for the defenders of science to invoke philosophy of science when it suits them (e.g., their much-loved principle of falsifiability comes directly from the philosopher Karl Popper) and to dismiss it as "arcane" and "remote" when it does not. However noble the motivation, bad philosophy makes for bad law.

The victory in the Arkansas case was hollow, for it was achieved only at the expense of perpetuating and canonizing a false stereotype of what science is and how it works. If it goes unchallenged by the scientific community, it will raise grave doubts about that community's intellectual integrity. No one familiar with the issues can really believe that anything important was settled through anachronistic efforts to revive a variety of discredited criteria for distinguishing between the scientific and the non-scientific. Fifty years ago, Clarence Darrow asked, *à propos* the Scopes trial, "Isn't it difficult to realize that a trial of this kind is possible in the twentieth century in the United States of America?" We can raise that question anew, with the added irony that, this time, the pro-science forces are defending a philosophy of science which is, in its way, every bit as outmoded as the "science" of the creationists.

■ | Notes

1. *McLean v. Arkansas Board of Education*, 529 F. Supp. 1255 (E.D. Ark. 1982). For the text of the law, the decision, and essays by participants in the trial, see 7 *Science, Technology, and Human Values* 40 (Summer 1982), and also *Creationism, Science, and the Law* [The Arkansas Case, ed. Marcel C. La Follette (Cambridge, Mass.: MIT Press, 1983).]
2. "The Creationist Threat: Science Finally Awakens," *The Skeptical Inquirer* 3 (Spring 1982): 2-5.

Response to the Commentary: Pro Judice

As always, my friend Larry Laudan writes in an entertaining and provocative manner, but, in his complaint against Judge William Overton's ruling in *McLean v. Arkansas*,¹ Laudan is hopelessly wide of the mark. Laudan's outrage centers on the criteria for the demarcation of science which Judge Overton adopted, and the judge's conclusion that, evaluated by these criteria, creation-science fails as science. I shall respond directly to this concern—after making three preliminary remarks.

First, although Judge Overton does not need defense from me or anyone else, as one who participated in the Arkansas trial, I must go on record as saying that I was enormously impressed by his handling of the case. His written judgment is a first-class piece of reasoning. With cause, many have criticized the State of Arkansas for passing the "Creation-Science Act," but we should not ignore that, to the state's credit, Judge Overton was born, raised, and educated in Arkansas.

Second, Judge Overton, like everyone else, was fully aware that proof that something is not science is not the same as proof that it is religion. The issue of what constitutes science arose because the creationists claim that their ideas qualify as genuine science rather than as fundamentalist religion. The attorneys developing the American Civil Liberties Union (ACLU) case believed it important to show that creation-science is not genuine science. Of course, this demonstration does raise the question of what creation-science really is. The plaintiffs claimed that creation-science always was (and still is) religion. The plaintiffs' lawyers went beyond the negative argument (against science) to make the positive case (for religion). They provided considerable evidence for the religious nature of creation-science, including such things as the creationists' explicit reliance on the Bible in their various writings. Such arguments seem about as

strong as one could wish, and they were duly noted by Judge Overton and used in support of his ruling. It seems a little unfair, in the context, therefore, to accuse him of "specious" argumentation. He did not adopt the naïve dichotomy of "science or religion but nothing else."

Third, whatever the merits of the plaintiffs' case, the kinds of conclusions and strategies apparently favored by Laudan are simply not strong enough for legal purposes. His strategy would require arguing that creation-science is weak science and therefore ought not to be taught:

The core issue is not whether Creationism satisfies some undemanding and highly controversial definitions of what is scientific; the real question is whether the existing evidence provides stronger arguments for evolutionary theory than for Creationism. Once that question is settled, we will know what belongs in the classroom and what does not.²

Unfortunately, the U.S. Constitution does not bar the teaching of weak science. What it bars (through the Establishment Clause of the First Amendment) is the teaching of religion. The plaintiffs' tactic was to show that creation-science is less than weak or bad science. It is not science at all.

Turning now to the main issue, I see three questions that must be addressed. Using the five criteria listed by Judge Overton, can one distinguish science from non-science? Assuming a positive answer to the first question, does creation-science fail as genuine science when it is judged by these criteria? And, assuming a positive answer to the second, does the Opinion in *McLean* make this case?

The first question has certainly tied philosophers of science in knots in recent years. Simple criteria that supposedly give a clear answer to every case—for example, Karl Popper's single stipulation of falsifiability³—will not do. Nevertheless, although there may be many grey areas, white does seem to be white and black does seem to be black. Less metaphorically, something like psychoanalytic theory may or may not be science, but there do appear to be clear-cut cases of real science and of real non-science. For instance, an explanation of the fact that my son has blue eyes, given that both parents have blue eyes, done in terms of dominant and recessive genes and with an appeal to Mendel's first law, is scientific. The Catholic doctrine of transubstantiation (i.e., that in the Mass the bread and wine turn into the body and blood of Christ) is not scientific.

Furthermore, the five cited criteria of demarcation do a good job of distinguishing the Mendelian example from the Catholic example. Law and explanation through law come into the first example. They do not enter the second. We can test the first example, rejecting it if necessary. In this sense, it is tentative, in that something empirical might change our minds. The case of transubstantiation is different. God may have His own

laws, but neither scientist nor priest can tell us about those which turn bread and wine into flesh and blood. There is no explanation through law. No empirical evidence is pertinent to the miracle. Nor would the believer be swayed by any empirical facts. Microscopic examination of the Host is considered irrelevant. In this sense, the doctrine is certainly not tentative.

One pair of examples certainly do not make for a definitive case, but at least they do suggest that Judge Overton's criteria are not quite as irrelevant as Laudan's critique implies. What about the types of objections (to the criteria) that Laudan does or could make? As far as the use of law is concerned, he might complain that scientists themselves have certainly not always been that particular about reference to law. For instance, consider the following claim by Charles Lyell in his *Principles of Geology* (1830/3): "We are not, however, contending that a real departure from the antecedent course of physical events cannot be traced in the introduction of man."⁴ All scholars agree that in this statement Lyell was going beyond law. The coming of man required special divine intervention. Yet, surely the *Principles* as a whole qualify as a contribution to science.

Two replies are open: either one agrees that the case of Lyell shows that science has sometimes mingled law with non-law; or one argues that Lyell (and others) mingled science and non-science (specifically, religion at this point). My inclination is to argue the latter. Insofar as Lyell acted as scientist, he appealed only to law. A century and a half ago, people were not as conscientious as today about separating science and religion. However, even if one argues the former alternative—that some science has allowed place for non-lawbound events—this hardly makes Laudan's case. Science, like most human cultural phenomena, has evolved. What was allowable in the early nineteenth century is not necessarily allowable in the late twentieth century. Specifically, science today does not break with law. And this is what counts for us. We want criteria of science for today, not for yesterday. (Before I am accused of making my case by fiat, let me challenge Laudan to find one point within the modern geological theory of plate tectonics where appeal is made to miracles, that is, to breaks with law. Of course, saying that science appeals to law is not asserting that we know all of the laws. But, who said that we did? Not Judge Overton in his Opinion.)

What about the criterion of tentativeness, which involves a willingness to test and reject if necessary? Laudan objects that real science is hardly all that tentative: "[H]istorical and sociological researches on science strongly suggest that the scientists of any epoch likewise regard some of their beliefs as so fundamental as not to be open to repudiation or negotiation."⁵

It cannot be denied that scientists do sometimes—frequently—hang

on to their views, even if not everything meshes precisely with the real world. Nevertheless, such tenacity can be exaggerated. Scientists, even Newtonians, have been known to change their minds. Although I would not want to say that the empirical evidence is all-decisive, it plays a major role in such mind changes. As an example, consider a major revolution of our own time, namely that which occurred in geology. When I was an undergraduate in 1960, students were taught that continents do not move. Ten years later, they were told that they do move. Where is the dogmatism here? Furthermore, it was the new empirical evidence—e.g., about the nature of the sea-bed—which persuaded geologists. In short, although science may not be as open-minded as Karl Popper thinks it is, it is not as close-minded as, say, Thomas Kuhn⁶ thinks it is.

Let me move on to the second and third questions, the status of creation-science and Judge Overton's treatment of the problem. The slightest acquaintance with the creation-science literature and Creationism movement shows that creation-science fails abysmally as science. Consider the following passage, written by one of the leading creationists, Duane T. Gish, in *Evolution: The Fossils Say No!*:

CREATION. By creation we mean the bringing into being by a supernatural Creator of the basic kinds of plants and animals by the process of sudden, or fiat, creation.

We do not know how the Creator created, what processes He used, *for He used processes which are not now operating anywhere in the natural universe.* This is why we refer to creation as Special Creation. We cannot discover by scientific investigations anything about the creative processes used by the Creator.⁷

The following similar passage was written by Henry M. Morris, who is considered to be the founder of the creation-science movement:

... it is ... quite impossible to determine anything about Creation through a study of present processes, because present processes are not created in character. If man wishes to know anything about Creation (the time of Creation, the duration of Creation, the order of Creation, the methods of Creation, or anything else) his sole source of true information is that of divine revelation. God was there when it happened. We were not there ... therefore, we are completely limited to what God has seen fit to tell us, and this information is in His written Word. This is our textbook on the science of Creation!⁸

By their own words, therefore, creation-scientists admit that they appeal to phenomena not covered or explicable by any laws that humans can grasp

as laws. It is not simply that the pertinent laws are not yet known. Creative processes stand outside law as humans know it (or could know it) on Earth—at least there is no way that scientists can know laws breaking (or transcending) Mendel's laws through observation and experiment. Even if God did use His own laws, they are necessarily veiled from us forever in this life, because Genesis says nothing of them.

Furthermore, there is nothing tentative or empirically checkable about the central claims of creation-science. Creationists admit as much when they join the Creation Research Society (the leading organization of the movement). As a condition of membership applicants must sign a document specifying that they now believe and will continue to believe:

(1) The Bible is the written Word of God, and because we believe it to be inspired throughout, all of its assertions are historically and scientifically true in all of the original autographs. To the student of nature, this means that the account of origins in Genesis is a factual presentation of simple historical truths. (2) All basic types of living things, including man, were made by direct creative acts of God during Creation Week as described in Genesis. Whatever biological changes have occurred since Creation have accomplished only changes within the original created kinds. (3) The great Flood described in Genesis, commonly referred to as the Noachian Deluge, was an historical event, worldwide in its extent and effect. (4) Finally, we are an organization of Christian men of science, who accept Jesus Christ as our Lord and Savior. The account of the special creation of Adam and Eve as one man and one woman, and their subsequent fall into sin, is the basis for our belief in the necessity of a Savior for all mankind. Therefore, salvation can come only thru accepting Jesus Christ as our Savior."

It is difficult to imagine evolutionists signing a comparable statement, that they will never deviate from the literal text of Charles Darwin's *On the Origin of Species*. The non-scientific nature of creation-science is evident for all to see, as is also its religious nature. Moreover, the quotes I have used above were all used by Judge Overton, in the *McLean* Opinion, to make exactly the points I have just made. Creation-science is not genuine science, and Judge Overton showed this.

Finally, what about Laudan's claim that some parts of creation-science (e.g., claims about the Flood) are falsifiable and that other parts (e.g., about the originally created "kinds") are revisable? Such parts are not falsifiable or revisable in a way indicative of genuine science. Creation-science is not like physics, which exists as part of humanity's common cultural heritage and domain. It exists solely in the imaginations and writing of a relatively small group of people. Their publications (and stated intentions) show that, for example, there is no way they will relinquish

belief in the Flood, whatever the evidence.¹⁰ In this sense, their doctrines are truly unfalsifiable.

Furthermore, any revisions are not genuine revisions, but exploitations of the gross ambiguities in the creationists' own position. In the matter of origins, for example, some elasticity could be perceived in the creationist position, given the conflicting claims that the possibility of (degenerative) change within the originally created "kinds." Unfortunately, any open-mindedness soon proves illusory; for creationists have no real idea about what God is supposed to have created in the beginning, except that man was a separate species. They rely solely on the Book of Genesis:

And God said, Let the waters bring forth abundantly the moving creature that hath life, and the fowl that may fly above the earth in the open firmament of heaven.

And God created great whales, and every living creature that moveth, which the waters brought forth abundantly, after their kind, and every winged fowl after his kind: and God saw that it was good.

And God blessed them, saying Be fruitful, and multiply, and fill the waters in the seas, and let fowl multiply in the earth.

And the evening and the morning were the fifth day.

And God said, Let the earth bring forth the living creature after his kind, cattle, and creeping thing, and beast of the earth after his kind: and it was so.

And God made the beast of the earth after his kind, and cattle after their kind, and everything that creepeth upon the earth after his kind: and God saw that it was good.¹¹

But the *definition* of "kind," what it really is, leaves creationists as mystified as it does evolutionists. For example, creationist Duane Gish makes this statement on the subject:

[W]e have defined a basic kind as including all of those variants which have been derived from a single stock. . . . We cannot always be sure, however, what constitutes a separate kind. The division into kinds is easier the more the divergence observed. It is obvious, for example, that among invertebrates the protozoa, sponges, jellyfish, worms, snails, trilobites, lobsters, and bees are all different kinds. Among the vertebrates, the fishes, amphibians, reptiles, birds, and mammals are obviously different basic kinds.

Among the reptiles, the turtles, crocodiles, dinosaurs, pterosaurs (flying reptiles), and ichthyosaurs (aquatic reptiles) would be placed in different kinds.

Each one of these major groups of reptiles could be further subdivided into the basic kinds within each.

Within the mammalian class, duck-billed platypus, bats, hedgehogs, rats, rabbits, dogs, cats, lemurs, monkeys, apes, and men are easily assignable to different basic kinds. Among the apes, the gibbons, orangutans, chimpanzees, and gorillas would each be included in a different basic kind.¹²

Apparently, a "kind" can be anything from humans (one species) to trilobites (literally thousands of species). The term is flabby to the point of inconsistency. Because humans are mammals, if one claims (as creationists do) that evolution can occur within but not across kinds, then humans could have evolved from common mammalian stock—but because humans themselves are kinds such evolution is impossible.

In brief, there is no true resemblance between the creationists' treatment of their concept of "kind" and the openness expected of scientists. Nothing can be said in favor of creation-science or its inventors. Overton's judgment emerges unscathed by Laudan's complaints.

■ | Notes

1. For the text of Judge Overton's Opinion, see 7 *Science, Technology, and Human Values* 40 (Summer 1982): 28–42; and *Creationism, Science, and the Law* [Cambridge, Mass.: MIT Press, 1983].
2. Larry Laudan, "Commentary: Science at the Bar—Causes for Concern," [p. 52 this volume].
3. Karl Popper, *The Logic of Scientific Discovery* (London: Hutchinson, 1959).
4. Charles Lyell, *Principles of Geology*, Volume 1 (London: John Murray, 1830), p. 162.
5. Laudan, *op. cit.*, [p. 50 this volume].
6. Thomas Kuhn, *The Structure of Scientific Revolutions* (Chicago, IL: University of Chicago Press, 1962).
7. Duane Gish, *Evolution: The Fossils Say No!*, 3rd edition (San Diego, CA: Creation-Life Publishers, 1979), p. 40 (his italics).
8. Henry M. Morris, *Studies in the Bible and Science* (Philadelphia, PA: Presbyterian and Reformed Publishing Company, 1966), p. 114.
9. Application form for the Creation Research Society, reprinted in Plaintiffs' trial briefs, *McLean v. Arkansas* (1981).

10. See, for instance, Henry M. Morris, *Scientific Creationism* (San Diego, CA: Creation-Life Publishers, 1974); and my own detailed discussion in Michael Ruse, *Darwinism Defended: A Guide to the Evolution Controversies* (Reading, MA: Addison-Wesley, 1982).
11. Genesis, Book I, Verses 20-25.
12. Gish, *op. cit.*, pp. 34-35.

1 | COMMENTARY

- 1.1 Popper's Demarcation Criterion 63
 - Falsifiability* 64
 - Popper and the Theory of Evolution* 65
- 1.2 Kuhn's Criticisms of Popper 66
 - Normal Science and Puzzle Solving* 67
 - Scientific Revolutions* 68
- 1.3 Lakatos and Scientific Research Programmes 69
 - Why All Theories Are Unprovable* 69
 - Why All Theories Are Improbable* 69
 - The Methodology of Scientific Research Programmes* 71
- 1.4 Thagard on Why Astrology Is a Pseudoscience 72
 - Thagard's Definition of Pseudoscience* 73
 - Thagard's Later Thoughts about Pseudoscience* 73
- 1.5 Creation-Science and the Arkansas Trial 74
 - Judge Overton's Opinion* 75
 - Ruse on the Status of Creation-Science* 76
 - Laudan's Criticisms of Ruse* 77
- 1.6 Summary 77

I | COMMENTARY

1.1 | Popper's Demarcation Criterion

In "Science: Conjectures and Refutations," Sir Karl Popper explains how he came to formulate his falsifiability criterion for the scientific status of a theory. He recognized that it was not enough to use the so-called empirical (or inductive) method of generalizing from observation and experience, for by this standard astrology might well qualify as genuine science. So why, Popper wondered, were Freudian psychoanalysis, Adlerian "individual psychology," and the Marxist theory of history more like astrology than astronomy, more like myth than science?

His answer came from noting that, while proponents of these disciplines found confirming evidence for their theories at every turn, they made no predictions that could be disconfirmed by evidence. With deliberate irony, Popper describes "the incessant stream of confirmations, of observations which 'verified' the theories in question" (5). Moreover, it seemed to Popper as though just about anything, even apparent counter-evidence, could be explained in Freudian or Adlerian or Marxist terms. In marked contrast to this were certain features characterizing one of the most important physical theories of this century. Popper recounts how impressed he was by the bold prediction of the bending of starlight near the surface of the sun made by Einstein's general theory of relativity. This prediction was verified by two astronomical expeditions to observe the total solar eclipse of 29 May 1919—one to Brazil, the other to the west coast of Africa—organized by the British cosmologist Sir Arthur Eddington. Photographic plates produced during these expeditions revealed that starlight was indeed deflected by the sun by an amount very close to Einstein's prediction of 1.75 seconds of arc. This crucial observation led to the overthrow of Newton's theory of gravity by Einstein's general theory of relativity.¹

Unlike Marx's theory of history and Adler's theory of the inferiority complex, Einstein's theory ran a serious risk of refutation by predicting the result of an observational test *before* the test was made. Popper sees this possibility of refutation by observation and experiment as the hallmark of genuine science. Agreement with known facts, or the ability to explain known facts, is not enough to make a theory scientific. Whereas the Marxists and Adlerians saw confirmation of their theories everywhere and recognized nothing that their theories could not explain, Einstein's theory is refutable because, by its very nature, it is incompatible with certain possible results of observation—it is open to falsifying tests. Popper insists that in order to be scientific a theory must take a risk by predicting something

new. Thus, Popper advocates falsifiability (testability), not verifiability (confirmability), as the demarcation criterion for distinguishing science from pseudoscience.

FALSIFIABILITY

Despite its simplicity and initial plausibility, there is much that is unclear and controversial about Popper's demarcation criterion. Part of the uncertainty arises because Popper shifts back and forth between two different notions—between *falsifiability* as a logical property of statements (requiring that scientific statements logically imply at least one testable prediction) and *falsifiability* as a term prescribing how scientists should act. According to Popper, scientists should test their theories by trying to refute them; when a prediction disagrees with observation and experiment, they should abandon their theories as refuted. Falsifiability in this second, prescriptive sense implies falsifiability in the first sense, for it is only by making testable predictions that a theory—made up of scientific statements—can be refuted. But the implication does not hold in the other direction. It is perfectly possible for a theory such as Marxism to imply at least one testable prediction (say, that all socialist revolutions will occur among the proletariat of industrialized capitalist nations) and yet, when the prediction turns out to be false (because the Russian and Chinese revolutions occurred in societies that were preindustrial and feudal), the adherents of the theory refuse to regard it as refuted and strive to explain away the anomaly. Thus, a theory that is scientific in Popper's (first) logical sense might be judged pseudoscientific in Popper's (second) methodological sense because of the behavior of its proponents.

Many philosophers have criticized the prescriptive, methodological aspect of Popper's demarcation criterion. They argue that abandoning a theory the instant it makes a false prediction would rule out too much good science. (This criticism, made by Kuhn and Lakatos, among others, will be discussed later.) Some philosophers also object to the first sense of Popper's falsifiability criterion, that falsifiability is a logical property of scientific statements, on the grounds that it is too weak. Take any statement, however implausible or crazy it may sound, and conjoin it with a respectable scientific theory. The crazy statement, *C*, might be the claim that aliens visited the earth during the Pleistocene era and removed all traces of their visit before departing. Although *C* is not a tautology, it makes no testable predictions. The respectable scientific theory, *T*, could be from any field whatever—geology, chemistry, physics, or astronomy. The conjunction, (*T* & *C*), makes lots of testable predictions since its logical consequences include all the predictions made by *T* alone. Thus, (*T* & *C*) satisfies Popper's falsifiability criterion. The moral is clear: having testable consequences is a very weak requirement. At best, perhaps, it is a necessary condition for genuine science, and many statements that satisfy

it are not part of science. Thus, presumably, Popper was not claiming that all falsifiable statements are scientific; he was merely claiming that in order to be scientific, a statement must be falsifiable.²

POPPER AND THE THEORY OF EVOLUTION

Popper claimed for several decades that the principle of natural selection in Darwin's theory of evolution fails to satisfy his falsifiability criterion (that is, falsifiability as a logical property of statements) and thus, in some important sense, that it is not scientific but "metaphysical." Popper recanted this belief, appropriately enough, when he delivered the first Darwin Lecture at Darwin College, Cambridge University, in 1977.³ This is important for three reasons. First, it illustrates how difficult it can be to decide whether or not a component of a scientific theory is falsifiable. Second, it illustrates the complexity of Popper's position, since Popper never condemned the whole of Darwin's theory as a pseudoscience even when he judged that an important part of that theory could not be falsified. Third, it sheds some light—if just a little—on the position of creationists who, much to Popper's dismay, have appealed to Popper's (pre-1977) writings for support in their crusade against the theory of evolution.

Before his recantation, Popper expressed reservations about Darwin's theory by saying that "Darwinism is not a testable scientific theory, but a *metaphysical research programme*—a possible framework for testable scientific theories."⁴ What Popper meant by this is that Darwin's theory, when expressed in very general terms as a group of claims about heredity, random mutation, and differential survival, does not make any predictions about which species (or indeed whether any species) will evolve. Popper thought that prediction and explanation seem to occur because we forget that adaptation or fitness is implicitly defined in terms of survival. Thus, while it may seem as if we have explained why a particular species now thrives by saying that it adapted to its environment, Popper judged this to be no explanation at all. Rather, he said that the claim that a species now living has adapted to its environment is "almost tautological," that is, true by definition.⁵

Popper's charge that the phrase "the survival of the fittest" is tantamount to a tautology (that to survive is to be fittest) has been repeated by creationists such as Henry Morris, who have then denied that evolutionary theory as whole is either empirical or testable.⁶ But even if a theory includes some elements that are true by definition or untestable for some other reason, it hardly follows that the theory as a whole or specific versions of it are untestable. Indeed, Popper regarded Darwinism as similar in this regard to atomism and field theory. In his view, these are all metaphysical generalizations that make no predictions and hence are untestable. Nonetheless, they are of great scientific value because they give rise to specific theories that *are* testable and have been tested. So when Popper judges a

proposition to be unfalsifiable and "metaphysical," he is not claiming that the proposition has no scientific value, nor is he asserting that any theory associated with it is pseudoscientific.

But is "the survival of the fittest" a tautology as has been charged? One problem in assessing this accusation is that "the survival of the fittest" is a phrase, not a proposition, and only propositions can be tautologies.⁷ What is needed is a precise statement of the allegedly tautologous proposition. Is the proposition in question a definition of *fitness* (or *relative adaptiveness*) in terms of the probability of reproductive success? Or is it the historical claim that the traits in current populations are the result of natural selection (i.e., selection of the fittest ancestral variants)? In his Darwin Lecture, Popper opted for the latter and then noted, correctly, that it is an empirical matter whether natural selection or some other mechanism (such as genetic drift) is responsible for the traits we now find in a population of organisms. Thus, Popper conceded that the principle of natural selection is falsifiable and testable.

Before concluding this section, there is one further small matter concerning Darwin's theory and Popper's criterion of falsifiability. Sometimes the claim is made (often, but not always, by creationists) that Darwin's theory (and, presumably, other sciences such as paleontology, geology, and cosmology) are unscientific because they are, at least in part, historical. Evolutionary theory, we are told, makes claims about historical events, many of which occurred before the advent of any human observers on this planet. Historical events are unique and unrepeatable. Therefore, critics conclude, Darwin's theory cannot be tested or refuted.⁸ This argument, as Popper himself has emphasized, is invalid: its conclusion does not follow from its premises.⁹ Claims about historical events, even events that occurred millions of years ago can be tested (and thus, in principle, refuted) by using them to make predictions about the evidence we should find now if the historical claims are true: cometary collisions with the earth leave craters and abnormally high concentrations of iridium in the surrounding rocks; animals and plants leave fossils; the "big bang" still resonates in the form of background microwave radiation in space.

1.2 | Kuhn's Criticisms of Popper

One of the many people who have challenged Popper's appeal to falsifiability as a demarcation criterion is Thomas Kuhn. In his book, *The Structure of Scientific Revolutions* (1962), Kuhn insisted that if we are to arrive at an adequate characterization of science, close attention must be paid to its history: a proper philosophy of science should reflect the history of science. On this account, philosophy of science ought to describe the way scientists actually behave and the way that science has evolved over time.

Doing so shows that not all scientific activity is of the same kind. Kuhn thinks that scientific activity falls into two distinct types: normal science and extraordinary (or revolutionary) science.

NORMAL SCIENCE AND PUZZLE SOLVING

During periods of normal science, scientists take for granted the major theories of their day and content themselves with what Kuhn calls puzzle solving. In some respects, the puzzle-solving aspect of normal science is like trying to do the exercises at the back of a physics or chemistry textbook. The aim of practicing scientists is not to call into question Newtonian mechanics or the laws of thermodynamics, but rather to see whether they can solve problems by using these accepted theories in conjunction with other assumptions and models. Just as failure to get the right answer to an exercise is regarded as a failure of the student, not of the theory, so, too, failure to solve a puzzle during a period of normal science is considered the fault of the scientist using the theory, not the fault of the theory itself. Only very rarely, during periods of extraordinary science, do scientists deliberately question the received theories of their day and attempt to refute them. Typically such periods of extraordinary science arise because of repeated failures to solve puzzles. If a theory is refuted, then it must be replaced by another theory that is at least as general in scope. Science, like nature, abhors a vacuum: scientists will give up a global theory only when they have an even better theory to adopt in its place. When such a replacement occurs, we have a scientific revolution. (For a much fuller discussion of Kuhn's views on scientific revolutions, see chapter 2, "Rationality, Objectivity, and Values in Science," below.)

Kuhn agrees with Popper and many other philosophers of science that astrology is a pseudoscience. In this, as in many other cases, Popper's criterion of demarcation (severity of testing) leads to the same verdict as Kuhn's criterion (puzzle solving). But Kuhn rejects Popper's demarcation criterion and with it Popper's explanation of why astrology is pseudoscientific. Popper insists that by formulating their accounts in suitably vague terms, astrologers are able to "explain away anything that might have been a refutation of the theory" (8). For Popper, this emphasis on confirmation and avoidance of testability or falsification marks the difference between pseudoscience and science. Kuhn's account of why astrology is a pseudoscience is quite different from Popper's. Kuhn points out that astrology was finally abandoned by scientists around the middle of the seventeenth century, mainly as a consequence of the Copernican revolution. But throughout its history, astrology was notoriously unreliable and its predictions often failed. Interestingly, these frequent failures were never given as a reason for thinking that astrology is false until after astrology had been abandoned. During its heyday, astrology was regarded rather as medicine and meteorology once were—as an imprecise study of an enor-

mously complex subject. In any case, astrology can scarcely be reckoned a nonscience simply because it made predictions that turned out to be false. Still, Kuhn insists, astrology never was a science. Although astrologers use rules of thumb to cast horoscopes, astrology has no central theory and no puzzle-solving tradition of a sort characterizing normal science. Thus astrology was, and remains, at best a craft and not a science.

SCIENTIFIC REVOLUTIONS

In "Logic of Discovery or Psychology of Research?" Kuhn gives several examples of scientific revolutions: the overthrow of Newton's theory of gravity by Einstein's general theory of relativity; the replacement of the phlogiston theory by Lavoisier's new chemical theory (in which the addition of oxygen, not the release of phlogiston, is responsible for the burning of metals in air); the experimental confirmation of Lee and Yang's theory that the weak interaction—the nuclear process responsible for the release of electrons during radioactive decay—does not conserve parity. One of Kuhn's main criticisms of Popper is that sincere attempts to refute theories are quite rare in science. Such attempts are usually confined to the periods of extraordinary science that immediately precede scientific revolutions. Thus, according to Kuhn, Popper's falsifiability account of science fails to describe normal science. If falsifiability were the criterion marking off science from pseudoscience, then genuine science as it is done most of the time, being normal and not extraordinary, would be improperly classified as pseudoscientific.

As we have seen, Kuhn rejects Popper's falsifiability criterion as an account of normal science; but how well does it fit those episodes of extraordinary science (scientific revolutions) in which large-scale theories are refuted and replaced? According to Kuhn, another flaw in Popper's historically insensitive treatment is that in some scientific revolutions—Kuhn gives the Copernican revolution as an example—the old theory (Ptolemy's geocentric theory) was replaced by the new theory (Copernicus's heliocentric theory) *before* the old theory was refuted. For example, Galileo's telescopic observations of the phases of Venus, the moons of Jupiter, and the motion of sunspots were made at least sixty years after the publication of Copernicus's *De Revolutionibus* (1543) and only after Galileo had become a convinced Copernican. Arguably, Ptolemy's theory (in which the earth is stationary at the center of the universe) was decisively refuted only when Newton's theory of mechanics and gravity was accepted. (Newton's *Principia* was published in 1687.) Newton's theory showed that it was physically impossible for the entire heavens to rotate around the earth's north-south axis. When Copernicus proposed his new theory, most astronomers thought that the Ptolemaic theory could solve all its problems by adjusting a few parameters. Hardly anyone thought that Ptolemy's theory had been severely tested and found irreparably wanting. Here again,

Kuhn argues, Popper's account of science does not fit the history of science. There is more to science and being scientific than falsifiability and testing.

1.3 | Lakatos and Scientific Research Programmes

In "Science and Pseudoscience," Imre Lakatos notes that genuine *scientia* (knowledge) cannot be marked off from impostors simply in terms of the number of people who believe it or how strongly they believe it. The worst of pseudoscience has, in the past, commanded dogged assent from large numbers of intellectuals. Nor can we rest a criterion of demarcation on the commonplace assertion that genuine science is supported by the observable facts. For, Lakatos asks, how could this criterion be justified? Like Kuhn and Popper, Lakatos agrees that no scientific theory can be deduced from observational and experimental facts. When scientists such as Newton and Ampère claimed that their theories were not *hypotheses* but proven *truths* because they were deduced from experiments and observations, they were simply wrong.

WHY ALL THEORIES ARE UNPROVABLE

We can appreciate Lakatos's point by considering a single example: Newton's theory of gravitation. Newton's theory says that *every* particle of matter in the universe attracts every other particle with a force according to an inverse square law. Newton's theory is a universal generalization that applies to every particle of matter, anywhere in the universe, at any time. But however numerous they might be, our observations of planets, falling bodies, and projectiles concern only a finite number of bodies during finite amounts of time. So the scope of Newton's theory vastly exceeds the scope of the evidence. It is possible that all our observations are correct, and yet Newton's theory is false because some bodies not yet observed violate the inverse square law. Since "All *F*s are *G*" cannot be deduced from "Some *F*s are *G*," it cannot be true that Newton's theory can be proven by logically deducing it from the evidence. As Lakatos points out, this prevents us from claiming that scientific theories, unlike pseudoscientific theories, can be proven from observational facts. The truth is that no theory can be deduced from such facts. *All* theories are unprovable, scientific and unscientific alike.

WHY ALL THEORIES ARE IMPROBABLE

While conceding that scientific theories cannot be proven, most people still believe that theories can be made more probable by evidence. Lakatos

follows Popper in denying that any theory can be made probable by any amount of evidence. Popper's argument for this controversial claim rests on the analysis of the objective probability of statements given by inductive logicians.¹⁰ Consider a card randomly drawn from a standard deck of fifty-two cards. What is the probability that the card selected is the ten of hearts? Obviously, the answer is $1/52$. There are fifty-two possibilities, each of which is equally likely and only one of which would render true the statement "This card is the ten of hearts." Now consider a scientific theory that, like Newton's theory of gravitation, is universal. The number of things to which Newton's theory applies is, presumably, infinite. Imagine that we name each of these things by numbering them 1, 2, 3, . . . , n , There are infinitely many ways the world could be, each equally probable.

1 obeys Newton's theory, but none of the others do.

1 and 2 obey Newton's theory, but none of the others do.

1, 2, and 3 obey Newton's theory, but none of the others do.

All bodies (1, 2, 3, . . . , n , . . .) obey Newton's theory.

Since these possibilities are infinite in number, and each of them has the same probability, the probability of any one of them must be 0.¹¹ But only one, the last one, represents the way the world would be if Newton's theory were true. So the probability of Newton's theory (and any other universal generalization) must be 0.

Now one might think that, even if the initial probability of a theory must be 0, the probability of the theory when it has been confirmed by evidence will be greater than 0. As it turns out, the probability calculus denies this. Let our theory be T , and let our evidence for T be E . We are interested in $P(T/E)$, the probability of T given our evidence E . Bayes's theorem (which follows logically from the axioms of the probability calculus) tells us that this probability is:

$$P(T/E) = \frac{P(E/T) \times P(T)}{P(E)}$$

If the initial probability of T —that is, $P(T)$ —is 0, then $P(T/E)$ must also be 0.¹² Thus, no theory can increase in objective probability, regardless of the amount of evidence for it. For this reason, Lakatos joins Popper in regarding all theories, whether scientific or not, as equally unprovable and equally improbable.

THE METHODOLOGY OF SCIENTIFIC RESEARCH PROGRAMMES

The failure to specify demarcation criteria along the intuitively attractive lines of "whatever is proved or made probable by evidence" might suggest returning to the Popperian model. But like Kuhn (and Thagard after him), Lakatos rejects Popper's falsifiability criterion as a solution to the demarcation problem. Scientists rarely specify in advance of observation and experiment those results that, if found, would refute their theories. At best, such results would be regarded as anomalous or recalcitrant, not as genuine refutations. Even when they are first proposed, some theories are (or are thought to be) inconsistent with the known data. Newton's gravitational theory is a good example. By his own admission, Newton was unable to reconcile his theory with the known orbit of the earth's moon. (This anomaly was later cleared up by Alexis Clairault who found a mistake in Newton's calculations.¹³) But Newton did not immediately abandon his theory as refuted. Later, after the discovery of the planet Uranus (by William Herschel in 1781), it was noted that Uranus did not move precisely as Newton's theory predicted. Again, scientists did not abandon the inverse square law; rather, they postulated another planet, as yet unobserved, which was perturbing the orbit of Uranus. This hypothetical new planet was eventually discovered and given the name Neptune.¹⁴

In order to make sense of the ways in which scientists protect their theories from refutation, Lakatos proposes that scientific theories be regarded as having three components: a hard core, a protective belt, and a positive heuristic. The hard core of Newton's theory consists of his three laws of motion plus the inverse square law of gravitational attraction. These are basic postulates that scientists were extremely reluctant to give up. The protective belt consists of many auxiliary hypotheses such as assumptions about the number and the masses of the planets. The positive heuristic tells scientists how to solve problems using the theory and how to respond to anomalies by revising the protective belt. Lakatos proposes that we stop thinking of scientific theories as frozen in time but instead regard theories as historically extended scientific research programmes. The Newtonian research programme covered several centuries. For much of its history it was progressive. Why? Because in dealing with anomalies and other problems, the Newtonian programme continued to predict novel facts.

According to Lakatos, Popper is wrong in thinking that a crucial experiment can (or should) instantly refute a theory. As Kuhn has shown, the actual history of science teaches us otherwise: genuine scientific progress (as opposed to degenerating science or pseudoscience) is not simply a matter of one theory remaining unrefuted while others are falsified. But Lakatos is equally critical of Kuhn for suggesting that scientific revolutions are largely irrational affairs, dependent on a kind of group psychology. Were Kuhn right, there would be no objective way of marking off scientific progress from scientific regress or decay. Instead, Lakatos suggests that

scientific change occurs as the result of competition between rival research programmes. If one programme is progressive (because it continues to predict novel facts), and if its rival is degenerating, then most scientists will, rationally, switch their allegiance. In this way, progressive research programmes replace degenerating ones.

1.4 | Thagard on Why Astrology Is a Pseudoscience

Paul Thagard takes up Lakatos's notion of scientific theories as research programmes, and develops it into an explicit criterion for demarcating science from pseudoscience. In "Why Astrology Is a Pseudoscience," Thagard surveys several different proposals for a demarcation criterion that would explain why astrology is a pseudoscience and finds each of them deficient. In light of the alchemical and occult origins of chemistry and medicine, one cannot uncritically cite astrology's origin in magic as what makes it a pseudoscience. (Indeed, this merely postpones the question, Why is magic not itself a genuine science?) Nor can the supposed immunity from testing, verification, or falsification be what makes astrology a pseudoscience. As Thagard notes, some astrological claims (about the influence of planetary positions at the time of one's birth on one's personality and future career, for example) are testable. Moreover, Thagard agrees with Kuhn and Lakatos that abandoning a theory the moment one of its predictions failed would be irrational. Many of our best scientific theories have been modified in the light of failed predictions and recalcitrant observations. Hasty rejection would nip too many good theories in the bud, before they had the chance to grow and blossom. (The ambiguity of falsification and the Duhem-Quine thesis are discussed further in chapter 3, "The Duhem-Quine Thesis and Underdetermination.")

Contrary to Kuhn, Thagard claims that modern astrology does indeed present a number of unsolved problems (such as accommodating the precession of the equinoxes and planets that were discovered many centuries after Ptolemy's death). This undercuts the Kuhnian proposal that astrology fails as a science simply because it is not a paradigm-dominated discipline of problem solving. Against Lakatos, Thagard suggests that lack of progress is not by itself a sufficient condition of pseudoscience, since it might be nonprogressive only in periods when it faces no progressive competitors. Despite these differences, however, Thagard agrees with Kuhn that judgments about the scientific status of a theory or discipline must involve both a social and a historical dimension, and he agrees with Lakatos that progress is necessary for genuine science.

THAGARD'S DEFINITION OF PSEUDOSCIENCE

In light of his criticisms of Kuhn and Lakatos, Thagard proposes two conditions that are necessary and sufficient for a theory or discipline to be pseudoscientific.

A theory or discipline which purports to be scientific is *pseudoscientific* if and only if:

- 1 it has been less progressive than alternative theories over a long period of time, and faces many unsolved problems; but
- 2 the community of practitioners makes little attempt to develop the theory towards solutions of the problems, shows no concern for attempts to evaluate the theory in relation to others, and is selective in considering confirmations and disconfirmations. (33)

According to these conditions, astrology is pseudoscientific in part because it has not changed much since the time of Ptolemy. Unsolved problems have accumulated, and as Thagard notes, we now have (since the nineteenth century) psychological theories that do a better job of explaining and predicting human behavior. Despite this competition from psychology, astrologers have shown little interest in improving their theory or in evaluating it with respect to rivals.

Thagard concludes by isolating a number of interesting logical (and, to some, startling) consequences of his demarcation criterion. One might view the acceptability of these consequences as a measure of the plausibility of his proposal. First, some current fads, such as pyramidology and biorhythms, would not be considered pseudosciences because, at the moment, they lack serious competitors. Second, a theory can be scientific at one time and pseudoscientific at a later time; being scientific is not an unchanging property of a theory. Third, Thagard concludes that astrology used to be a genuine science but became pseudoscientific only when modern psychology arose in the late nineteenth century. If this is correct, then those scientists (the vast majority) who rejected astrology as pseudoscientific in the eighteenth century were being irrational.

THAGARD'S LATER THOUGHTS ABOUT PSEUDOSCIENCE

Because of objections to his demarcation principle for pseudoscience, especially the objection that nothing can be a pseudoscience unless it has competitors, Thagard has changed his views. In his book *Computational Philosophy of Science* (1988) he gives up trying to provide necessary and sufficient conditions for pseudoscience. Instead, he offers contrasting profiles of genuine science and pseudoscience. Relative progressiveness and

a concern with confirmation and disconfirmation are still presented as hallmarks of science (and their absence is still associated with pseudoscience), but Thagard no longer claims that science must always possess these features or that pseudoscience must necessarily lack them. Thagard also introduces two new criteria for pseudoscience. One of these criteria is that pseudoscientific theories are often highly complex and riddled with ad hoc hypotheses. This provides some grounds for judging a doctrine pseudoscientific on its content, even if it currently has no scientific competitors.

Thagard's second new criterion concerns the sort of reasoning employed by many practitioners of pseudoscience, such as astrologers, namely, reasoning based on resemblances. Instead of testing causal claims by looking for statistical correlations, pseudoscientists are often content to rest their beliefs on superficial analogies. Traditional astrology is full of this sort of "resemblance thinking." For example, the planet Mars often has a reddish appearance, and so astrologers associate it with blood, war, and aggression. From this they conclude that Mars causes (or, at least, has a tendency to cause) aggressive personalities in people born at the appropriate time. In a similar way believers in folk medicine recommend turmeric as a treatment for jaundice and powdered rhinoceros horn as a cure for impotence.

As Thagard recognizes, not all pseudosciences employ resemblance thinking, and some pseudosciences employ reasoning based on statistical correlations that mimics, to some extent, reasoning found in the genuine sciences. Proponents of biorhythms, for example, rest much of their case on alleged correlations as do Velikovsky and von Daniken when they appeal to common elements in ancient myths to support their astronomical theories. Thus, in Thagard's revised account of pseudoscience, none of the elements mentioned—using resemblance thinking, refusing to seek confirmations and disconfirmations, ignoring alternative theories, trafficking in ad hoc hypotheses, sticking with theories that fail to progress—is a necessary feature of pseudoscience, and genuine sciences might, from time to time, share one or two of these features. But, Thagard claims, pseudosciences usually have most of these features and genuine sciences nearly always lack most of them. Thus, the difference between science and pseudoscience is a matter of degree rather than kind, although Thagard remains convinced that the difference of degree is usually large and obvious.

1.5 | Creation-Science and the Arkansas Trial

The search for demarcation criteria is not simply a curiosity to entertain armchair intellectuals or a pastime for students in philosophy of science.

Consider, for instance, the 1982 case of *McLean v. Arkansas Board of Education*.¹⁵ At issue in the case was the constitutionality of Arkansas Act 590, which required teachers to give "balanced treatment" to both evolutionary theory and creationism in the biology classes taught in public schools. Act 590 describes "evolution-science" and "creation-science" as competing scientific models of the origin of species and offers the following definition for creation-science:

"Creation-science" means the scientific evidences [sic] for creation and inferences from those scientific evidences. Creation-science includes the scientific evidences and related inferences that indicate: (1) Sudden creation of the universe, energy, and life from nothing; (2) The insufficiency of mutation and natural selection in bringing about development of all living kinds from a single organism; (3) Changes only within fixed limits of originally created kinds of plants and animals; (4) Separate ancestry for man and apes; (5) Explanation of the earth's geology by catastrophism, including the occurrence of a worldwide flood; and (6) A relatively recent inception of the earth and living kinds.¹⁶

JUDGE OVERTON'S OPINION

The task of the presiding judge, William Overton, was to decide whether Act 590 violates the Constitution of the United States. He reasoned that Act 590 is consistent with the Constitution only if the act satisfies the Establishment Clause of the First Amendment, which says that "Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof." The Supreme Court of the United States has for many years applied the articles of the Bill of Rights (the first ten amendments to the Constitution) not only to federal legislation but also to the laws passed by individual states. The Supreme Court's interpretation of the Establishment Clause has evolved into a three-part test for the constitutionality of any legislation involving religion. It was this three-part test that Judge Overton applied to Arkansas's Act 590. Failing any one of these three parts is sufficient to render a piece of legislation unconstitutional. Here is the test:

First, the statute must have a secular purpose; second, its principal or primary effect must be one that neither advances nor inhibits religion . . . ; finally, the statute must not foster "an excessive government entanglement with religion."¹⁷

Judge Overton thought it clear that Act 590 was passed by the Arkansas General Assembly with the specific intention of advancing religion, and that fact alone—the lack of a secular purpose—would suffice to invalidate the statute. But Judge Overton wanted to show that Act 590 also fails the

second and third parts of the three-part test. In order to show that Act 590 fails the second part, it is necessary to show that the statute either advances or inhibits religion as its "principal or primary effect." To accomplish this, Judge Overton thought it necessary to establish that creation-science is not a genuine science. For, as he argued (at the end of part IV(D) of his Opinion), "Since creation-science is not science, the conclusion is inescapable that the *only* real effect of Act 590 is the advancement of religion."¹⁸

Thus Judge Overton entered the philosophical debate over the criteria for genuine science. He sought guidance from expert witnesses, especially from a philosopher of biology, Michael Ruse.¹⁹ It was primarily Ruse who developed the five characteristics that Overton lists as essential (necessary conditions) for genuine science:

- 1 it is guided by natural law;
- 2 it has to be explanatory by reference to natural law;
- 3 it is testable against the empirical world;
- 4 its conclusions are tentative, i.e., they are not necessarily the final word; and
- 5 it is falsifiable.

RUSE ON THE STATUS OF CREATION-SCIENCE

In "Creation-Science Is Not Science" Ruse defends the five items on Overton's list and argues that creation-science satisfies none of them. Ruse sees an intimate connection between items (1) and (2) on the list: it is only because scientific theories posit natural laws that the theories are able to explain; genuinely to explain something is to show why, given the relevant circumstances, it had to happen, and that requires an appeal to laws. (For more on explanation and laws, see chapter 6, "Models of Explanation," and chapter 7, "Laws of Nature.") Since creation-science posits acts of creation that are miraculous and unlawful, Ruse concludes that it is not scientific. He also points out that creation-scientists make few if any testable predictions. Most of the time, creationists content themselves with describing the evidence in ways that are consistent with their doctrines. For example, creationists regard the common pattern of bones in the forelimbs of humans, bats, whales, and other mammals as an instance of God's design plan for mammals, but they offer no reason why this particular pattern exists rather than some other pattern or several different patterns. Evolutionists follow Darwin in explaining the pattern as the result of common descent: because all mammals have descended from a common ancestor, they share a common anatomical structure. Ruse concludes his case against creation-science by noting that most creationist research aims at trying to find flaws in evolutionary theory rather than making testable

predictions based on the creationists' own theory. Modern creationists are dogmatic (not tentative) about their fundamental beliefs and show little or no interest in trying to falsify them.

LAUDAN'S CRITICISMS OF RUSE

In his "Commentary: Science at the Bar—Causes for Concern," Larry Laudan chastizes Ruse for perpetuating a view of science that, he claims, both of them know to be false. Laudan denies that philosophers of science would accept any list of characteristics as capturing the essence of science. Why, for example, should explanation by means of laws be regarded as a necessary condition for a theory to be scientific? Many theories begin by describing a new phenomenon, and only later, if at all, explain the phenomenon in a lawlike way. For example, Galileo discovered that all bodies released near the surface of the earth fall with the same acceleration but offered no explanation for this. Similarly, Newton claimed to have "deduced" the universal law of gravitation "from the phenomena" but accepted action at a distance as ultimately unexplainable.²⁰ Indeed, if one accepts the deductive-nomological model of explanation, according to which scientific explanations are deductive arguments with at least one statement of a law in their premises (see chapter 6, "Models of Explanation"), then in any such explanation there will remain, at least provisionally, something that is not explained, namely the premises that do the explaining. So Laudan rejects item (2) from the Ruse-Overton list as too strong.

Laudan also criticizes items (3) and (5) as being too weak, since, he argues, they are all too easily satisfied. Any theory, even a theory like creation-science that posits a divine creator, implies something about the observable world. For example, many creationists claim that all living things were created at the same time fewer than 50,000 years ago and that a worldwide flood caused many of the geological features now observed on the earth. As Laudan sees it, the flaw with creationism is not that such claims are untestable or unfalsifiable but rather that they have been tested and falsified.²¹ Ruse responds to these and other criticisms by Laudan in the final piece in this chapter, "Response to the Commentary: *Pro Judice*."²²

1.6 | Summary

In this chapter, we have explored a number of attempts to demarcate science from pseudoscience. But the results have been curiously inconclusive. Most scientists and philosophers of science readily agree that such things as pyramidology and creation-science are not genuine sciences, but

there is no consensus on why this is so. Like obscenity, most people are able to recognize pseudoscience when they encounter it but find it much harder to explain why what they have encountered is pseudoscientific. The stress on explanation is important here. What we are seeking, as philosophers of science, is not just a handy way of detecting pseudoscience (on the basis, say, of a majority vote of the National Academy of Sciences) but a philosophically informative account of what makes a discipline genuinely scientific.

Despite the defects of his own demarcation criterion—falsifiability—Popper deserves credit for disposing of one tempting answer to the demarcation problem. No appeal to confirming evidence, by itself, is going to distinguish genuine science from its counterfeit. Inventing an elaborate hypothesis that is consistent with the known facts is just too easy. Popper's fruitful idea was to seek the demarcation between science and pseudoscience, not in confirmation, but in falsification. The hallmark of true science is its willingness to make testable predictions. If the predictions fail, then the theory should be abandoned as false. Unfortunately, Popper's simple idea does not work. As Lakatos and Thagard explain, falsifiability is both too weak and too strong. It is too weak because it would allow as scientific any number of claims that are testable in principle but that are, by no stretch of the imagination, scientific. It is too strong because it would rule out as unscientific many of the best theories in the history of science. Few scientists give up their theories simply because they have come into conflict with observation and experiment. Instead, they either look for a flaw in the data, or they modify their theories. The rejection of a theory simply because it disagrees with the facts (or what are taken to be facts) is the exception rather than the rule in science.

In differing ways, Kuhn, Lakatos, and Thagard each propose that there is a historical and a social dimension to judgments concerning the scientific status of a theory. All three insist that when we ask of a theory "Is it genuinely scientific?" it is a mistake to look at the theory as if it were a snapshot, caught at an instant of time. Rather, they argue, we have to consider how the theory has developed, especially how the theory has been modified to deal with new problems and recalcitrant data. For Kuhn, this means seeing the theory as part of a larger whole—what Kuhn calls a *paradigm*. (For more on Kuhn's notion of a paradigm see chapter 2, "Rationality, Objectivity, and Values in Science.") Thagard adopts Lakatos's notion of a scientific research programme in order to define a demarcation criterion. On this approach, roughly speaking, a theory is pseudoscientific if the research programme with which it is associated has been less progressive over time than has its rivals. As suggested in our discussion of Thagard's proposal, this comparative-progress definition of pseudoscience has a number of startling consequences. For example, some modern fads, such as pyramidology, might fail to qualify as pseudosciences simply because, at the moment, they lack competitors. Because of these defects,

Thagard has avoided giving necessary and sufficient conditions for pseudoscience in his more recent writings.

As the debate between Ruse and Laudan concerning the status of creation-science makes clear, judgments about pseudosciences often depend on detailed considerations of the nature of law, explanation, confirmation, and falsification. It is highly unlikely that any simple-minded, one- or two-sentence definition of science will yield a plausible demarcation criterion that we can use to label and condemn as pseudoscientific those theories (and their advocates) that fail to meet the standards of good science. Ultimately, discriminating between science and its counterfeit depends on a detailed understanding of how science works. Despite the variety and complexity of the many different theories and activities that are, by common consent, genuinely scientific, are there general principles concerning explanation, confirmation, testing, and the like that these theories and activities share? In the rest of our book, some important attempts to answer this question will be explained and evaluated. Thus, what follows can be seen as an attempt to answer the questions left unanswered in this first chapter.

Notes

1. Newton's theory also predicts the bending of starlight if light rays are regarded as a stream of particles traveling at the speed of light. Because inertial mass is exactly equal to gravitational mass, the orbit of any object moving around the sun depends only on the velocity of the moving object, not on its mass. (The same thing is true of bodies near the surface of the earth. If you throw two objects with the same velocity in the same direction, then they will follow the same path regardless of their mass.) Thus, we do not have to know the mass of the light particles in order to calculate how they will move when close to the sun. But, Newton's theory predicts an amount of bending which is only half of that predicted by Einstein. The difference arises because Einstein's theory entails that the gravitational field close to the sun is slightly stronger than in Newton's theory. Thus, it is not that Einstein's theory predicted a kind of effect, the bending of starlight, that Newton's theory did not. Rather, both theories gave competing predictions of its magnitude, and Einstein's prediction was more nearly right. Interestingly, observations made during some later eclipses (1929, 1947) found deviations that were higher than those predicted by Einstein. But more recent observations are in closer agreement with Einstein's theory, and none of the observations agrees with Newton's. (The issue of whether theories such as Newton's theory of gravity can be conclusively refuted is discussed in chapter 3, "The Duhem-Quine Thesis and Underdetermination.") Eddington's role in this episode is controversial because he threw out as biased one set of observations that agreed with Newton's prediction. For details about the difficulties of making the eclipse observations and competing interpretations of Eddington's behavior, see John Earman and Clark Glymour, "Relativity and Eclipses: The British Eclipse Expeditions of 1919 and

Their Predecessors," *Historical Studies in the Physical Sciences* 11 (1980): 49–85; Deborah Mayo, "Novel Evidence and Severe Tests," *Philosophy of Science* 58 (1991): 523–52; and Harry Collins and Trevor Pinch, *The Golem: What Everyone Should Know about Science* (Cambridge: Cambridge University Press, 1993).

2. It is tempting to try to rule out cases such as (T & C) by requiring not merely that the theory as a whole make testable predictions but that each individual component of the theory also make testable predictions. See chapter 3 for a discussion of whether any significant scientific theory could meet this additional requirement.

3. See Karl R. Popper, "Natural Selection and the Emergence of Mind," *Dialectica* 32 (1978): 339–55.

4. Karl R. Popper, "Autobiography of Karl Popper," in *The Philosophy of Karl Popper*, ed. Paul A. Schilpp (La Salle, Ill.: Open Court, 1974), 1: 134.

5. *The Philosophy of Karl Popper*, 1: 137. For excellent discussions of the tautology problem, see chapter 2 of Elliott Sober's *The Nature of Selection* (Chicago, Ill.: University of Chicago Press, 1984), and chapter 4, "The Structure of the Theory of Natural Selection," in Robert N. Brandon's *Adaptation and Environment* (Princeton, N.J.: Princeton University Press, 1990).

6. Thus, somewhat paradoxically given the Arkansas trial (discussed later in this commentary), one of the themes in creationist literature is that creationism and evolutionary theory are both equally unscientific because neither makes testable predictions. Needless to say, this position is hard to reconcile with another creationist theme, namely, that evolutionary theory has been significantly disconfirmed by a variety of evidence.

7. Another problem is that tautologies, strictly speaking, are propositions that are true solely in virtue of their logical form. Presumably, the issue is not whether some biological statement is a tautology but whether it is analytic. Analytic statements, on one characterization of analyticity, are statements that are true solely in virtue of the meanings of the words and symbols used to express them. In chapter 3, there is an extended discussion of Quine's thesis that no line can be drawn, even in principle, between statements that are analytic and those that are not. If Quine is right, then the charge of being tautologous evaporates.

8. See, for example, the authors quoted in Beverly Halstead, "Popper: Good Philosophy, Bad Science?" *New Scientist* (17 July 1980): 215–17.

9. Karl R. Popper, "Letter on Evolution," *New Scientist* (21 August 1980): 611.

10. The argument that follows is a simplified version of the one given in Appendix "vii of Karl R. Popper, *The Logic of Scientific Discovery* (New York: Basic Books, 1959), 363–77.

11. The possible hypotheses enumerated in the text—exactly one particle obeys Newton's theory, exactly two particles obey Newton's theory, etc.—are exclusive: if any one of them is true, then all the others must be false. If each hypothesis has the same prior probability, p , and there are n of them, then the probability that at least one of the hypotheses is true is $n \times p$. (See axiom 3, the special addition rule, in "Bayes's Theorem and the Axioms of Probability Theory" in the commentary on chapter 5.) Since $n \times p$ is a probability, it cannot be greater than

1. So, if n is infinite, p cannot be finite. Thus, p must be 0. The derivation of this result depends on assuming that each hypothesis has the same prior probability, something that Bayesians deny. For this and other Bayesian criticisms of Popper's argument, see Colin Howson, "Must the Logical Probability of Laws Be Zero?" *British Journal for the Philosophy of Science* 35 (1973): 153–63.

12. For a fuller discussion of this and other applications of Bayes's theorem to issues in confirmation, see chapter 5, "Confirmation and Relevance: Bayesian Approaches."

13. As Newton realized, the main irregularities in the motion of the moon are due to the attraction of the sun. The force exerted on the moon by the sun is a rather large fraction ($1/89$ at new and full moon) of the force exerted by the earth. As in all such three-body problems, no exact solution of Newton's equations is possible. Because the moon is close to the earth, even small perturbations are easily observed. This requires the calculations to be extended down to very small terms. Initially, Clairault's calculations yielded a rate of precession of the moon's apogee of 20 degrees per year, only half the real amount. At first, Clairault speculated that Newton's inverse-square law gravitational formula was incorrect for small distances and should be supplemented by an extra term varying as the inverse fourth power of the distance. But on extending his calculations to include higher order terms that had been neglected in his original approximation, Clairault found that his first result was doubled. So Newton's theory was vindicated. For more on the problem of the moon, see Anton Pannekoek, *A History of Astronomy*, rpt. (1951; New York: Dover Publications, 1990) ch. 30.

14. See Morton Grosser, *The Discovery of Neptune* (Cambridge, Mass.: Harvard University Press, 1962). For a lively criticism of the oft-repeated claim (by Popper, Lakatos, and others) that Newton's theory was *prima facie* falsified by the discovery of perturbations in the orbit of Uranus, see Greg Bamford, "Popper and His Commentators on the Discovery of Neptune: A Close Shave for the Law of Gravitation?" *Studies in History and Philosophy of Science* 27 (1996): 207–32.

15. Judge Overton's opinion in this case is reprinted in *Science* 215 (1982): 934–43, in *Science, Technology, and Human Values* 7 No. 40 (1982): 28–42, and in Michael Ruse, ed., *But Is It Science?* (Buffalo, N.Y.: Prometheus Books, 1988).

16. *Balanced Treatment for Creation-Science and Evolution-Science Act* (1981), 73d General Assembly, State of Arkansas, Act 590 sec. 4; reprinted in *Science, Technology, and Human Values* 7 No. 40 (1982): 11.

17. William R. Overton, "Opinion in *McLean v. Arkansas*," *Science, Technology, and Human Values* 7 No. 40 (1982): 29.

18. For a criticism of this inference and other aspects of Overton's opinion, see Philip L. Quinn, "The Philosopher of Science as Expert Witness," in *Science and Reality*, ed. J. T. Cushing, C. F. Delaney, and G. M. Gutting (Notre Dame, Ind.: University of Notre Dame Press, 1984), 32–53. As Quinn explains in a later article, he agrees with Overton's conclusion that Arkansas Act 590 is unconstitutional because it lacks a secular purpose. What he criticizes is Overton's attempt to show that Act 590 has the advancement of religion as its primary effect because, as it is alleged, creation-science fails each of the five conditions on Ruse's list deemed necessary for genuine science. Like Laudan, Quinn argues that each of Ruse's

conditions is either not necessary for genuine science (because some bona fide sciences lack it) or, when properly interpreted, is possessed by creation-science. According to Quinn, the proper thing to say about creation-science is not that we can show that it is not science but that, at best, it is dreadful science. See Philip L. Quinn, "Creationism, Methodology, and Politics," in Michael Ruse, ed., *But Is It Science?* 395–99.

19. For an entertaining account of Ruse's participation in the Arkansas trial and a transcript of his testimony, see Michael Ruse, ed., *But Is It Science?* 13–35, 287–306.

20. In the General Scholium of the *Principia*, added to the second edition of 1713, Newton wrote: "But hitherto I have not been able to discover the cause of those properties of gravity [i.e., the proportionality of gravitational force to the quantity of matter and its variation with the inverse square of distance], and I frame no hypotheses [*hypotheses non fingo*]; for whatever is not deduced from the phenomena is to be called an hypothesis; and hypotheses, whether metaphysical or physical, whether of occult qualities or mechanical, have no place in experimental philosophy. In this philosophy particular propositions are inferred from the phenomena, and afterwards rendered general by induction." Isaac Newton, *Philosophiæ naturalis principia mathematica*, vol. 2, trans. A. Motte, rev. F. Cajori (Berkeley: University of California Press, 1934), 547. Throughout the *Principia*, Newton argues that gravity cannot be explained by any mechanism acting by direct contact, as Descartes and the Cartesians had hypothesized. This seems to leave only two choices, both of which Newton entertained in his writings: either gravity is due to the direct action of God, or it is caused by an aether, itself composed of particles between which forces act at a distance across empty space. Many philosophers of science, notably Duhem and Popper, have criticized Newton's claim that his own theory can be "deduced from the phenomena." See "Duhem's Critique of Inductivism: The Attack on Newtonian Method," in the commentary on chapter 3.

21. Elsewhere, Laudan has argued that the wide diversity of scientific beliefs and activities and the failure of the attempts by Popper, Thagard, and others to solve the demarcation problem make it unlikely that we will ever find a demarcation criterion in the form of necessary conditions for genuine science. See Larry Laudan, "The Demise of the Demarcation Problem," in *Physics, Philosophy, and Psychoanalysis*, ed. R. S. Cohen and L. Laudan (Dordrecht, Netherlands: D. Reidel, 1983), 111–28.

22. The debate between Ruse and Laudan is continued in Larry Laudan, "More on Creationism," *Science, Technology, and Human Values* 8 (Winter 1983): 36–38, and Michael Ruse, "The Academic as Expert Witness," *Science, Technology, and Human Values* 11 (Spring 1986): 68–73. These and other relevant articles are conveniently reprinted in Michael Ruse, ed., *But Is It Science?*



2

Rationality, Objectivity, and Values in Science

INTRODUCTION

Much of the recent debate about the role of values in science and the nature of scientific objectivity has been crucially affected by the work of Thomas Kuhn. Clark Glymour has aptly described Kuhn's *The Structure of Scientific Revolutions* as "very likely the single most influential work on the philosophy of science that has been or will be written in this century."¹ It is certainly the most widely read. Since its publication in 1962, the University of Chicago Press has sold over one-and-a-half million copies of the English-language edition, and the book had been translated into at least nineteen languages. The frequency with which the term *paradigm* is bandied about in disciplines as diverse as literary criticism, art history, sociology of knowledge, and theology attest to the far-reaching character of Kuhn's work.

What made Kuhn's book so controversial² was its rejection of many of the ways of thinking about science that had become standard during the first half of the twentieth century. Instead of giving logical analyses of individual scientific theories or constructing formal models of concepts such as explanation and confirmation, Kuhn turned to psychology, sociology, and history in order to draw a picture of science that, he claimed, was far more faithful to the original than anything that philosophers of science had yet proposed. The first reading in this chapter, "The Nature and Necessity of Scientific Revolutions," is taken from the heart of Kuhn's path-breaking book. In it Kuhn explains why he thinks that reason and evidence can play only a limited role in determining the outcome of

scientific revolutions and why we must abandon the traditional assumption that science progresses by getting ever closer to the truth.

Soon after *The Structure of Scientific Revolutions* was published, Kuhn was criticized for portraying science, at least during scientific revolutions, as an irrational affair that is largely “a matter for mob psychology.”² Many philosophers of science also deplored Kuhn’s rejection of objective progress and realism in favor of relativism about truth and instrumentalism about theories. In response to these criticisms, Kuhn added a substantial postscript to *The Structure of Scientific Revolutions* in 1970 and delivered a lecture in 1973, “Objectivity, Value Judgment, and Theory Choice,” which is reprinted as the second reading in this chapter. In this widely cited paper, Kuhn tried to defuse his critics’ accusations of irrationalism and subjectivism by acknowledging that, during scientific revolutions, proponents of rival paradigms often share a number of important cognitive values. To the extent that arguments can be based on these shared values, scientific revolutions are rational. But Kuhn still insisted that there is no set of universal rules for choosing between rival theories, that cognitive values are ultimately a matter of subjective preference that transcends rationality, and that nonrational psychological and social factors must play a vital role in determining which theory wins the allegiance of the scientific community. Moreover, Kuhn continued to reject scientific realism, persisting in his view that scientific theories should be regarded as instruments for solving puzzles rather than as literal descriptions (or would-be descriptions) of reality.

Despite their influence on other academic disciplines, Kuhn’s views are primarily about science, and it is on the basis of their adequacy as a philosophically informative account of science that they must be assessed. The next two readings in this chapter (by Ernan McMullin and Larry Laudan) explain and evaluate Kuhn’s claims about the role of values in science, especially during those upheavals that constitute scientific revolutions. In “Rationality and Paradigm Change in Science,” McMullin takes Kuhn to task for denying that the notions of objective progress and truth are relevant to understanding scientific revolutions, illustrating his criticism of Kuhn’s instrumentalism with an analysis of the Copernican revolution in which the sun-centered astronomy of Copernicus replaced the earth-centered astronomy of Ptolemy. Laudan also attacks Kuhn, but from a different angle. In “Dissecting the Holist Picture of Scientific Change,” he accuses Kuhn of having adopted a false and misleading picture of scientific rationality, a picture that Laudan calls the *hierarchical model*. Laudan proposes his own *reticulation model* of scientific rationality and uses it to expose the flaws in Kuhn’s arguments for relativism. Like McMullin, Laudan thinks that Kuhn is wrong to deny that there can be rational debate about cognitive values, but unlike McMullin, Laudan does not think that it is necessary to embrace realism in order to make that debate intelligible.

In the remaining two readings, by Helen Longino and Kathleen Okruhlik respectively, the emphasis shifts from cognitive values in scientific revolutions to contextual values in normal science. In Longino's "Values and Objectivity," the main topic is the role that contextual values (in the form of background beliefs, possibly of a sexist or racist nature) play in everyday scientific assumptions about theories by shaping what scientists recognize as evidence for their theories. Two key questions are whether such beliefs could be eliminated from science, even in principle, and whether the presence of such beliefs prevents science from being objective. Longino defends a contextualist analysis of evidence and locates scientific objectivity, not in rules for choosing between theories, but in a social organization that permits diverse viewpoints and encourages criticism. In "Gender and the Biological Sciences," Okruhlik gives examples of androcentric bias in the life sciences to illustrate the ways in which contextual values can affect scientific judgment. She then compares and contrasts several different feminist critiques of science and defends a version of feminist empiricism.

Since all the readings in this chapter concern the role of values in the practice of science, it is helpful to consider what kinds of things values are and what sorts of values are relevant to the question of scientific objectivity and rationality. The important distinction between cognitive and contextual values is explored in the first section of the commentary that follows the readings. This section also includes a brief discussion of the value-neutrality thesis—that is, the thesis that only cognitive values should play a role in the decisions that scientists make about theories.

Notes

1. Clark Glymour, *Theory and Evidence* (Princeton, N.J.: Princeton University Press, 1980), 94. The quotation comes from the opening sentence of a chapter entitled "New Fuzziness and Old Problems." By attesting to Kuhn's influence on the discipline, Glymour is not endorsing the approach of Kuhn and his successors to the philosophy of science. Rather, he berates the "new fuzziness" for failing to solve the "old problems" of explaining what makes evidence relevant to theory and why evidence varies in its confirming power. These traditional questions of confirmation theory are addressed in chapters 4 and 5 below.
2. Imre Lakatos, "Falsification and the Methodology of Scientific Research Programmes," in *Criticism and the Growth of Knowledge*, ed. I. Lakatos and A. Musgrave (Cambridge: Cambridge University Press, 1970), 178. Kuhn replies to this and similar criticisms in the second reading in chapter 2, "Objectivity, Value Judgment, and Theory Choice."

THOMAS S. KUHN

The Nature and Necessity of Scientific Revolutions

... What are scientific revolutions, and what is their function in scientific development? ... Scientific revolutions are here taken to be those non-cumulative developmental episodes in which an older paradigm is replaced in whole or in part by an incompatible new one. There is more to be said, however, and an essential part of it can be introduced by asking one further question. Why should a change of paradigm be called a revolution? In the face of the vast and essential differences between political and scientific development, what parallelism can justify the metaphor that finds revolutions in both?

One aspect of the parallelism must already be apparent. Political revolutions are inaugurated by a growing sense, often restricted to a segment of the political community, that existing institutions have ceased adequately to meet the problems posed by an environment that they have in part created. In much the same way, scientific revolutions are inaugurated by a growing sense, again often restricted to a narrow subdivision of the scientific community, that an existing paradigm has ceased to function adequately in the exploration of an aspect of nature to which that paradigm itself had previously led the way. In both political and scientific development the sense of malfunction that can lead to crisis is prerequisite to revolution. Furthermore, though it admittedly strains the metaphor, that parallelism holds not only for the major paradigm changes, like those attributable to Copernicus and Lavoisier, but also for the far smaller ones associated with the assimilation of a new sort of phenomenon, like oxygen or X-rays. Scientific revolutions ... need seem revolutionary only to those whose paradigms are affected by them. To outsiders they may, like the Balkan revolutions of the early twentieth century, seem normal parts of the developmental process. Astronomers, for example, could accept X-rays

FROM Thomas S. Kuhn, *The Structure of Scientific Revolutions*, 2d ed. (Chicago: University of Chicago Press, 1970), 92-110.

as a mere addition to knowledge, for their paradigms were unaffected by the existence of the new radiation. But for men like Kelvin, Crookes, and Roentgen, whose research dealt with radiation theory or with cathode ray tubes, the emergence of X-rays necessarily violated one paradigm as it created another. That is why these rays could be discovered only through something's first going wrong with normal research.*

This genetic aspect of the parallel between political and scientific development should no longer be open to doubt. The parallel has, however, a second and more profound aspect upon which the significance of the first depends. Political revolutions aim to change political institutions in ways that those institutions themselves prohibit. Their success therefore necessitates the partial relinquishment of one set of institutions in favor of another, and in the interim, society is not fully governed by institutions at all. Initially it is crisis alone that attenuates the role of political institutions as we have already seen it attenuate the role of paradigms. In increasing numbers individuals become increasingly estranged from political life and behave more and more eccentrically within it. Then, as the crisis deepens, many of these individuals commit themselves to some concrete proposal for the reconstruction of society in a new institutional framework. At that point the society is divided into competing camps or parties, one seeking to defend the old institutional constellation, the others seeking to institute some new one. And, once that polarization has occurred, *political recourse fails*. Because they differ about the institutional matrix within which political change is to be achieved and evaluated, because they acknowledge no supra-institutional framework for the adjudication of revolutionary difference, the parties to a revolutionary conflict must finally resort to the techniques of mass persuasion, often including force. Though revolu-

* Kuhn discusses the discovery of X rays in chapter 6 of *The Structure of Scientific Revolutions* and in his paper, "The Historical Structure of Scientific Discovery," *Science* 136 (1962): 760-64, reprinted in T. S. Kuhn, *The Essential Tension* (Chicago, Ill.: University of Chicago Press, 1977), 165-77. X rays are produced in a high-tension vacuum tube when cathode rays hit the glass wall of the tube. X rays were discovered accidentally by Roentgen in 1895 during his investigation of cathode rays. Roentgen noticed that a radiation ~~detection screen~~ at some distance from his vacuum tube glowed when current was passing through the tube, even though the apparatus was shielded. After seven weeks of intense experimental work, Roentgen ruled out cathode rays as a cause of the glow and announced his discovery of a new form of radiation that could pass easily through matter. His announcement was initially greeted with skepticism and surprise; Kelvin, for example, at first thought it must be a hoax.

Kuhn regards Roentgen's discovery as paradigm breaking because it required physicists to revise the way they performed and interpreted their experiments with cathode ray tubes. It also inspired the search for other new forms of radiation such as gamma rays and the notorious "N rays." (N rays were ultimately proved to be spurious—see Irving J. Langmuir, "Pathological Science," *Physics Today* 42 (1989): 36-48, for a revealing account of how respectable scientists can believe they are detecting and measuring phenomena that turn out to be nonexistent.)

tions have had a vital role in the evolution of political institutions, that rôle depends upon their being partially extrapolitical or extrainstitutional events.

The remainder of this essay aims to demonstrate that the historical study of paradigm change reveals very similar characteristics in the evolution of the sciences.* Like the choice between competing political institutions, that between competing paradigms proves to be a choice between incompatible modes of community life. Because it has that character, the choice is not and cannot be determined merely by the evaluative procedures characteristic of normal science, for these depend in part upon a particular paradigm, and that paradigm is at issue. When paradigms enter, as they must, into a debate about paradigm choice, their role is necessarily circular. Each group uses its own paradigm to argue in that paradigm's defense.

The resulting circularity does not, of course, make the arguments wrong or even ineffectual. The man who premises a paradigm when arguing in its defense can nonetheless provide a clear exhibit of what scientific practice will be like for those who adopt the new view of nature. That exhibit can be immensely persuasive, often compellingly so. Yet, whatever its force, the status of the circular argument is only that of persuasion. It cannot be made logically or even probabilistically compelling for those who refuse to step into the circle. The premises and values shared by the two parties to a debate over paradigms are not sufficiently extensive for that. As in political revolutions, so in paradigm choice—there is no standard higher than the assent of the relevant community. To discover how scientific revolutions are effected, we shall therefore have to examine not only the impact of nature and of logic, but also the techniques of persuasive argumentation effective within the quite special groups that constitute the community of scientists.

To discover why this issue of paradigm choice can never be unequivocally settled by logic and experiment alone, we must shortly examine the nature of the differences that separate the proponents of a traditional paradigm from their revolutionary successors. That examination is the principal object of this section We have, however, already noted numerous examples of such differences, and no one will doubt that history can supply many others. What is more likely to be doubted than their existence—and what must therefore be considered first—is that such examples provide essential information about the nature of science. Granting that paradigm rejection has been a historic fact, does it illuminate more than human credulity and confusion? Are there intrinsic reasons why the

* By "the remainder of this essay," Kuhn means the final five chapters of *The Structure of Scientific Revolutions* of which the present selection (from chapter 9) is a crucial part.

assimilation of either a new sort of phenomenon or a new scientific theory must demand the rejection of an older paradigm?

First notice that if there are such reasons, they do not derive from the logical structure of scientific knowledge. In principle, a new phenomenon might emerge without reflecting destructively upon any part of past scientific practice. Though discovering life on the moon would today be destructive of existing paradigms (these tell us things about the moon that seem incompatible with life's existence there), discovering life in some less well-known part of the galaxy would not. By the same token, a new theory does not have to conflict with any of its predecessors. It might deal exclusively with phenomena not previously known, as the quantum theory deals (but, significantly, not exclusively) with subatomic phenomena unknown before the twentieth century. Or again, the new theory might be simply a higher level theory than those known before, one that linked together a whole group of lower level theories without substantially changing any. Today, the theory of energy conservation provides just such links between dynamics, chemistry, electricity, optics, thermal theory, and so on. Still other compatible relationships between old and new theories can be conceived. Any and all of them might be exemplified by the historical process through which science has developed. If they were, scientific development would be genuinely cumulative. New sorts of phenomena would simply disclose order in an aspect of nature where none had been seen before. In the evolution of science new knowledge would replace ignorance rather than replace knowledge of another and incompatible sort.

Of course, science (or some other enterprise, perhaps less effective) might have developed in that fully cumulative manner. Many people have believed that it did so, and most still seem to suppose that cumulation is at least the ideal that historical development would display if only it had not so often been distorted by human idiosyncrasy. There are important reasons for that belief. . . . Nevertheless, despite the immense plausibility of that ideal image, there is increasing reason to wonder whether it can possibly be an image of *science*. After the pre-paradigm period the assimilation of all new theories and of almost all new sorts of phenomena has in fact demanded the destruction of a prior paradigm and a consequent conflict between competing schools of scientific thought. Cumulative acquisition of unanticipated novelties proves to be an almost non-existent exception to the rule of scientific development. The man who takes historic fact seriously must suspect that science does not tend toward the ideal that our image of its cumulativeness has suggested. Perhaps it is another sort of enterprise.

If, however, resistant facts can carry us that far, then a second look at the ground we have already covered may suggest that cumulative acquisition of novelty is not only rare in fact but improbable in principle. Normal research, which *is* cumulative, owes its success to the ability of

scientists regularly to select problems that can be solved with conceptual and instrumental techniques close to those already in existence. (That is why an excessive concern with useful problems, regardless of their relation to existing knowledge and technique, can so easily inhibit scientific development.) The man who is striving to solve a problem defined by existing knowledge and technique is not, however, just looking around. He knows what he wants to achieve, and he designs his instruments and directs his thoughts accordingly. Unanticipated novelty, the new discovery, can emerge only to the extent that his anticipations about nature and his instruments prove wrong. Often the importance of the resulting discovery will itself be proportional to the extent and stubbornness of the anomaly that foreshadowed it. Obviously, then, there must be a conflict between the paradigm that discloses anomaly and the one that later renders the anomaly lawlike. . . . There is no other effective way in which discoveries might be generated.

The same argument applies even more clearly to the invention of new theories. There are, in principle, only three types of phenomena about which a new theory might be developed. The first consists of phenomena already well explained by existing paradigms, and these seldom provide either motive or point of departure for theory construction. When they do, . . . the theories that result are seldom accepted, because nature provides no ground for discrimination. A second class of phenomena consists of those whose nature is indicated by existing paradigms but whose details can be understood only through further theory articulation. These are the phenomena to which scientists direct their research much of the time, but that research aims at the articulation of existing paradigms rather than at the invention of new ones. Only when these attempts at articulation fail do scientists encounter the third type of phenomena, the recognized anomalies whose characteristic feature is their stubborn refusal to be assimilated to existing paradigms. This type alone gives rise to new theories. Paradigms provide all phenomena except anomalies with a theory-determined place in the scientist's field of vision.

But if new theories are called forth to resolve anomalies in the relation of an existing theory to nature, then the successful new theory must somewhere permit predictions that are different from those derived from its predecessor. That difference could not occur if the two were logically compatible. In the process of being assimilated, the second must displace the first. Even a theory like energy conservation, which today seems a logical superstructure that relates to nature only through independently established theories, did not develop historically without paradigm destruction. Instead, it emerged from a crisis in which an essential ingredient was the incompatibility between Newtonian dynamics and some recently formulated consequences of the caloric theory of heat. Only after the caloric theory had been rejected could energy conservation become part of science.¹ And only after it had been part of science for some time could it

come to seem a theory of a logically higher type, one not in conflict with its predecessors.* It is hard to see how new theories could arise without these destructive changes in beliefs about nature. Though logical inclusiveness remains a permissible view of the relation between successive scientific theories, it is a historical implausibility.

A century ago it would, I think, have been possible to let the case for the necessity of revolutions rest at this point. But today, unfortunately, that cannot be done because the view of the subject developed above cannot be maintained if the most prevalent contemporary interpretation of the nature and function of scientific theory is accepted. That interpretation, closely associated with early logical positivism and not categorically rejected by its successors, would restrict the range and meaning of an accepted theory so that it could not possibly conflict with any later theory that made predictions about some of the same natural phenomena. The best-known and the strongest case for this restricted conception of a scientific theory emerges in discussions of the relation between contemporary Einsteinian dynamics and the older dynamical equations that descend from Newton's *Principia*. From the viewpoint of this essay these two theories are fundamentally incompatible in the sense illustrated by the relation of Copernican to Ptolemaic astronomy: Einstein's theory can be accepted only with the recognition that Newton's was wrong. Today this remains a minority view.² We must therefore examine the most prevalent objections to it.

The gist of these objections can be developed as follows. Relativistic dynamics cannot have shown Newtonian dynamics to be wrong, for Newtonian dynamics is still used with great success by most engineers and, in selected applications, by many physicists. Furthermore, the propriety of this use of the older theory can be proved from the very theory that has, in other applications, replaced it. Einstein's theory can be used to show that predictions from Newton's equations will be as good as our measuring instruments in all applications that satisfy a small number of restrictive conditions. For example, if Newtonian theory is to provide a good ap-

* According to the caloric theory, heat is a conserved fluid: its total quantity remains constant in its interactions with matter. So, for example, during the operation of a heat engine, no heat is destroyed when work is produced. This contradicts the energy conservation principle, according to which heat must be converted into an equivalent amount of work when the engine operates. As Kuhn writes, acceptance of the conservation-of-energy principle required the rejection of the caloric theory of heat. This rejection was hard to achieve because of the many well-confirmed results in thermodynamics obtained by Sadi Carnot using the caloric theory. Thus, accepting the energy conservation principle required not only overturning the prevailing caloric paradigm but also rewriting the foundations of thermodynamics. See D. S. L. Cardwell, *From Watt to Clausius: The Rise of Thermodynamics in the Early Industrial Age* (Ithaca, N.Y.: Cornell University Press, 1971), for a fascinating account of this revolution in physics.

proximate solution, the relative velocities of the bodies considered must be small compared with the velocity of light. Subject to this condition and a few others, Newtonian theory seems to be derivable from Einsteinian, of which it is therefore a special case.*

But, the objection continues, no theory can possibly conflict with one of its special cases. If Einsteinian science seems to make Newtonian dynamics wrong, that is only because some Newtonians were so incautious as to claim that Newtonian theory yielded entirely precise results or that it was valid at very high relative velocities. Since they could not have had any evidence for such claims, they betrayed the standards of science when they made them. In so far as Newtonian theory was ever a truly scientific theory supported by valid evidence, it still is. Only extravagant claims for the theory—claims that were never properly parts of science—can have been shown by Einstein to be wrong. Purged of these merely human extravagances, Newtonian theory has never been challenged and cannot be.

Some variant of this argument is quite sufficient to make any theory ever used by a significant group of competent scientists immune to attack. The much-maligned phlogiston theory, for example, gave order to a large number of physical and chemical phenomena. It explained why bodies burned—they were rich in phlogiston—and why metals had so many more properties in common than did their ores. The metals were all compounded from different elementary earths combined with phlogiston, and the latter, common to all metals, produced common properties. In addition, the phlogiston theory accounted for a number of reactions in which acids were formed by the combustion of substances like carbon and sulphur. Also, it explained the decrease of volume when combustion occurs in a confined volume of air—the phlogiston released by combustion “spoils” the elasticity of the air that absorbed it, just as fire “spoils” the elasticity of a steel spring.³ If these were the only phenomena that the phlogiston theorists had claimed for their theory, that theory could never have been challenged. A similar argument will suffice for any theory that has ever been successfully applied to any range of phenomena at all.

But to save theories in this way, their range of application must be restricted to those phenomena and to that precision of observation with which the experimental evidence in hand already deals.⁴ Carried just a step further (and the step can scarcely be avoided once the first is taken),

* Attempts to understand scientific change by viewing the relation between earlier theories and their successors as a type of reduction are discussed in chapter 8. The sort of relation between Newton's theory and Einstein's that Kuhn is alluding to here is treated as an instance of a domain-preserving reduction in the selection by Thomas Nickles, “Two Concepts of Intertheoretic Reduction” included there. A sustained attack on the traditional account of reduction—including the derivability requirement discussed by Kuhn below—can be found in Paul Feyerabend, “How to Be a Good Empiricist” also included in chapter 8.

such a limitation prohibits the scientist from claiming to speak "scientifically" about any phenomenon not already observed. Even in its present form the restriction forbids the scientist to rely upon a theory in his own research whenever that research enters an area or seeks a degree of precision for which past practice with the theory offers no precedent. These prohibitions are logically unexceptionable. But the result of accepting them would be the end of the research through which science may develop further.

By now that point too is virtually a tautology. Without commitment to a paradigm there could be no normal science. Furthermore, that commitment must extend to areas and to degrees of precision for which there is no full precedent. If it did not, the paradigm could provide no puzzles that had not already been solved. Besides, it is not only normal science that depends upon commitment to a paradigm. If existing theory binds the scientist only with respect to existing applications, then there can be no surprises, anomalies, or crises. But these are just the signposts that point the way to extraordinary science. If positivistic restrictions on the range of a theory's legitimate applicability are taken literally, the mechanism that tells the scientific community what problems may lead to fundamental change must cease to function. And when that occurs, the community will inevitably return to something much like its pre-paradigm state, a condition in which all members practice science but in which their gross product scarcely resembles science at all. Is it really any wonder that the price of significant scientific advance is a commitment that runs the risk of being wrong?

More important, there is a revealing logical lacuna in the positivist's argument, one that will reintroduce us immediately to the nature of revolutionary change. Can Newtonian dynamics really be *derived* from relativistic dynamics? What would such a derivation look like? Imagine a set of statements, E_1, E_2, \dots, E_n , which together embody the laws of relativity theory. These statements contain variables and parameters representing spatial position, time, rest mass, etc. From them, together with the apparatus of logic and mathematics, is deducible a whole set of further statements including some that can be checked by observation. To prove the adequacy of Newtonian dynamics as a special case, we must add to the E_i 's additional statements, like $(v/c)^2 \ll 1$, restricting the range of the parameters and variables. This enlarged set of statements is then manipulated to yield a new set, N_1, N_2, \dots, N_m , which is identical in form with Newton's laws of motion, the law of gravity, and so on. Apparently Newtonian dynamics has been derived from Einsteinian, subject to a few limiting conditions.

Yet the derivation is spurious, at least to this point. Though the N_i 's are a special case of the laws of relativistic mechanics, they are not Newton's Laws. Or at least they are not unless those laws are reinterpreted in a way that would have been impossible until after Einstein's work. The

variables and parameters that in the Einsteinian E_i 's represented spatial position, time, mass, etc., still occur in the N_i 's; and they there still represent Einsteinian space, time, and mass. But the physical referents of these Einsteinian concepts are by no means identical with those of the Newtonian concepts that bear the same name. (Newtonian mass is conserved; Einsteinian is convertible with energy. Only at low relative velocities may the two be measured in the same way, and even then they must not be conceived to be the same.) Unless we change the definitions of the variables in the N_i 's, the statements we have derived are not Newtonian. If we do change them, we cannot properly be said to have *derived* Newton's Laws, at least not in any sense of "derive" now generally recognized. Our argument has, of course, explained why Newton's Laws ever seemed to work. In doing so it has justified, say, an automobile driver in acting as though he lived in a Newtonian universe. An argument of the same type is used to justify teaching earth-centered astronomy to surveyors. But the argument has still not done what it purported to do. It has not, that is, shown Newton's Laws to be a limiting case of Einstein's. For in the passage to the limit it is not only the forms of the laws that have changed. Simultaneously we have had to alter the fundamental structural elements of which the universe to which they apply is composed.

This need to change the meaning of established and familiar concepts is central to the revolutionary impact of Einstein's theory. Though subtler than the changes from geocentrism to heliocentrism, from phlogiston to oxygen, or from corpuscles to waves [as an account of the nature of light], the resulting conceptual transformation is no less decisively destructive of a previously established paradigm. We may even come to see it as a prototype for revolutionary reorientations in the sciences. Just because it did not involve the introduction of additional objects or concepts, the transition from Newtonian to Einsteinian mechanics illustrates with particular clarity the scientific revolution as a displacement of the conceptual network through which scientists view the world.

These remarks should suffice to show what might, in another philosophical climate, have been taken for granted. At least for scientists, most of the apparent differences between a discarded scientific theory and its successor are real. Though an out-of-date theory can always be viewed as a special case of its up-to-date successor, it must be transformed for the purpose. And the transformation is one that can be undertaken only with the advantages of hindsight, the explicit guidance of the more recent theory. Furthermore, even if that transformation were a legitimate device to employ in interpreting the older theory, the result of its application would be a theory so restricted that it could only restate what was already known. Because of its economy, that restatement would have utility, but it could not suffice for the guidance of research.

Let us, therefore, now take it for granted that the differences between successive paradigms are both necessary and irreconcilable. Can we then

say more explicitly what sorts of differences these are? The most apparent type has already been illustrated repeatedly. Successive paradigms tell us different things about the population of the universe and about that population's behavior. They differ, that is, about such questions as the existence of subatomic particles, the materiality of light, and the conservation of heat or of energy. These are the substantive differences between successive paradigms, and they require no further illustration. But paradigms differ in more than substance, for they are directed not only to nature but also back upon the science that produced them. They are the source of the methods, problem-field, and standards of solution accepted by any mature scientific community at any given time. As a result, the reception of a new paradigm often necessitates a redefinition of the corresponding science. Some old problems may be relegated to another science or declared entirely "unscientific." Others that were previously non-existent or trivial may, with a new paradigm, become the very archetypes of significant scientific achievement. And as the problems change, so, often, does the standard that distinguishes a real scientific solution from a mere metaphysical speculation, word game, or mathematical play. The normal-scientific tradition that emerges from a scientific revolution is not only incompatible but often actually incommensurable with that which has gone before.

The impact of Newton's work upon the normal seventeenth-century tradition of scientific practice provides a striking example of these subtler effects of paradigm shift. Before Newton was born the "new science" of the century had at last succeeded in rejecting Aristotelian and scholastic explanations expressed in terms of the essences of material bodies. To say that a stone fell because its "nature" drove it toward the center of the universe had been made to look a mere tautological word-play, something it had not previously been. Henceforth the entire flux of sensory appearances, including color, taste, and even weight, was to be explained in terms of the size, shape, position, and motion of the elementary corpuscles of base matter. The attribution of other qualities to the elementary atoms was a resort to the occult and therefore out of bounds for science. Molière caught the new spirit precisely when he ridiculed the doctor who explained opium's efficacy as a soporific by attributing to it a dormitive potency.* During the last half of the seventeenth century many scientists

* Molière satirizes *virtus dormitiva* (dormitive potency) as an explanation of opium's power to induce sleep in his last play, *Le malade imaginaire* (The imaginary invalid, 1673). As Kuhn suggests (with deliberate irony) in his next sentence, the purported explanations given by the mechanico-corpuscular philosophy could be just as superficial as those offered by the Aristotelians and scholastics. More importantly, as Kuhn notes in the following paragraphs, the *vis inertiae* (force of inertia) and gravitational action-at-a-distance that Newton attributed to matter bear more than a casual resemblance to the essences and powers of the Aristotelians. In this respect, Newton was more of an Aristotelian than his corpuscularian con-

preferred to say that the round shape of the opium particles enabled them to sooth the nerves about which they moved.⁵

In an earlier period explanations in terms of occult qualities had been an integral part of productive scientific work. Nevertheless, the seventeenth century's new commitment to mechanico-corpuscular explanation proved immensely fruitful for a number of sciences, ridding them of problems that had defied generally accepted solution and suggesting others to replace them. In dynamics, for example, Newton's three laws of motion are less a product of novel experiments than of the attempt to reinterpret well-known observations in terms of the motions and interactions of primary neutral corpuscles. Consider just one concrete illustration. Since neutral corpuscles could act on each other only by contact, the mechanico-corpuscular view of nature directed scientific attention to a brand-new subject of study, the alteration of particulate motions by collisions. Descartes announced the problem and provided its first putative solution. Huyghens, Wren, and Wallis carried it still further, partly by experimenting with colliding pendulum bobs, but mostly by applying previously well-known characteristics of motion to the new problem. And Newton embedded their results in his laws of motion. The equal "action" and "reaction" of the third law are the changes in quantity of motion experienced by the two parties to a collision. The same change of motion supplies the definition of dynamical force implicit in the second law. In this case, as in many others during the seventeenth century, the corpuscular paradigm bred both a new problem and a large part of that problem's solution.⁶

Yet, though much of Newton's work was directed to problems and embodied standards derived from the mechanico-corpuscular world view, the effect of the paradigm that resulted from his work was a further and partially destructive change in the problems and standards legitimate for science. Gravity, interpreted as an innate attraction between every pair of particles of matter, was an occult quality in the same sense as the scholastics' "tendency to fall" had been. Therefore, while the standards of corpuscularism remained in effect, the search for a mechanical explanation of gravity was one of the most challenging problems for those who accepted the *Principia* as paradigm. Newton devoted much attention to it and so did many of his eighteenth-century successors. The only apparent option was to reject Newton's theory for its failure to explain gravity, and that alternative, too, was widely adopted. Yet neither of these views ultimately triumphed. Unable either to practice science without the *Principia*

temporaries. Newton's achievement lay not in banishing essences and powers from science but in discovering the precise mathematical laws according to which they operate and being able to use those laws to make testable predictions. For further thoughts in this direction, see Rudolf Carnap, "The Value of Laws: Explanation and Prediction" in chapter 6 below.

or to make that work conform to the corpuscular standards of the seventeenth century, scientists gradually accepted the view that gravity was indeed innate. By the mid-eighteenth century that interpretation had been almost universally accepted, and the result was a genuine reversion (which is not the same as a retrogression) to a scholastic standard. Innate attractions and repulsions joined size, shape, position, and motion as physically irreducible primary properties of matter.⁷

The resulting change in the standards and problem-field of physical science was once again consequential. By the 1740's, for example, electricians could speak of the attractive "virtue" of the electric fluid without thereby inviting the ridicule that had greeted Molière's doctor a century before. As they did so, electrical phenomena increasingly displayed an order different from the one they had shown when viewed as the effects of a mechanical effluvium that could act only by contact. In particular, when electrical action-at-a-distance became a subject for study in its own right, the phenomenon we now call charging by induction could be recognized as one of its effects. Previously, when seen at all, it had been attributed to the direct action of electrical "atmospheres" or to the leakages inevitable in any electrical laboratory. The new view of inductive effects was, in turn, the key to Franklin's analysis of the Leyden jar and thus to the emergence of a new and Newtonian paradigm for electricity. Nor were dynamics and electricity the only scientific fields affected by the legitimization of the search for forces innate to matter. The large body of eighteenth-century literature on chemical affinities and replacement series also derives from this supramechanical aspect of Newtonianism. Chemists who believed in these differential attractions between the various chemical species set up previously unimagined experiments and searched for new sorts of reactions. Without the data and the chemical concepts developed in that process, the later work of Lavoisier and, more particularly, of Dalton would be incomprehensible.^{8*} Changes in the standards governing permissible problems, concepts, and explanations can transform a science. . . .

* The oxygen theory of Antoine Lavoisier (1743-94) owed much to experiments on calcination and the isolation of gases by Joseph Priestley, Carl Scheele, and Henry Cavendish, all of whom were proponents of the phlogiston theory. Similarly, the atomic theory of John Dalton (1766-1844) was indebted to the discovery of the law of equivalent proportions (the basis for assigning equivalent weights to chemical elements) and the law of constant proportions (that regardless of how a compound is made, it always contains the same ratio of elements by weight), discoveries made within the Newtonian paradigm referred to by Kuhn. The laws of equivalent and constant proportions led Dalton to formulate the law of multiple proportions (that when two elements can form more than one compound, the weights of one element that combine with a fixed weight of the other are always in a simple numerical ratio) and the confirmation of this law played a central role in Dalton's case for the atomic theory.

Other examples of these nonsubstantive differences between successive paradigms can be retrieved from the history of any science in almost any period of its development. For the moment let us be content with just two other and far briefer illustrations. Before the chemical revolution, one of the acknowledged tasks of chemistry was to account for the qualities of chemical substances and for the changes these qualities underwent during chemical reactions. With the aid of a small number of elementary "principles"—of which phlogiston was one—the chemist was to explain why some substances are acidic, others metalline, combustible, and so forth. Some success in this direction had been achieved. We have already noted that phlogiston explained why the metals were so much alike, and we could have developed a similar argument for the acids. Lavoisier's reform, however, ultimately did away with chemical "principles," and thus ended by depriving chemistry of some actual and much potential explanatory power. To compensate for this loss, a change in standards was required. During much of the nineteenth century failure to explain the qualities of compounds was no indictment of a chemical theory.⁹

Or again, Clerk Maxwell shared with other nineteenth-century proponents of the wave theory of light the conviction that light waves must be propagated through a material ether. Designing a mechanical medium to support such waves was a standard problem for many of his ablest contemporaries. His own theory, however, the electromagnetic theory of light, gave no account at all of a medium able to support light waves, and it clearly made such an account harder to provide than it had seemed before. Initially, Maxwell's theory was widely rejected for those reasons. But, like Newton's theory, Maxwell's proved difficult to dispense with, and as it achieved the status of a paradigm, the community's attitude toward it changed. In the early decades of the twentieth century Maxwell's insistence upon the existence of a mechanical ether looked more and more like lip service, which it emphatically had not been, and the attempts to design such an ethereal medium were abandoned. Scientists no longer thought it unscientific to speak of an electrical "displacement" without specifying what was being displaced.* The result, again, was a new set of

* Kuhn is referring to the difficulty of understanding what the term D represents physically in Maxwell's equations for the electromagnetic field in free space. Maxwell introduced the term D , calling it the displacement current by analogy with the ordinary electric current that flows when a wire is connected to the terminals of a battery. When a state of electric polarization is induced in a dielectric medium, there is a transient change in the electric field. This changing electric field acts just like an electric current in producing a magnetic field. When the medium is a real, physical substance (such as an insulator between the plates of a condenser), we can easily imagine that the displacement current arises because charged particles (electrons) are moved slightly in the direction of the applied electric field. But what happens in free space, where there is no physical substance and no charged particles? According to Maxwell's equations, variations in the displacement current give rise to a changing magnetic field that, in turn, induces an

problems and standards, one which, in the event, had much to do with the emergence of relativity theory.¹⁰

These characteristic shifts in the scientific community's conception of its legitimate problems and standards would have less significance to this essay's thesis if one could suppose that they always occurred from some methodologically lower to some higher type. In that case their effects, too, would seem cumulative. No wonder that some historians have argued that the history of science records a continuing increase in the maturity and refinement of man's conception of the nature of science.¹¹ Yet the case for cumulative development of science's problems and standards is even harder to make than the case for cumulation of theories. The attempt to explain gravity, though fruitfully abandoned by most eighteenth-century scientists, was not directed to an intrinsically illegitimate problem; the objections to innate forces were neither inherently unscientific nor metaphysical in some pejorative sense. There are no external standards to permit a judgment of that sort. What occurred was neither a decline nor a raising of standards, but simply a change demanded by the adoption of a new paradigm. Furthermore, that change has since been reversed and could be again. In the twentieth century Einstein succeeded in explaining gravitational attractions, and that explanation has returned science to a set of canons and problems that are, in this particular respect, more like those of Newton's predecessors than of his successors. Or again, the development of quantum mechanics has reversed the methodological prohibition that originated in the chemical revolution. Chemists now attempt, and with great success, to explain the color, state of aggregation, and other qualities of the substances used and produced in their laboratories. A similar reversal may even be underway in electromagnetic theory. Space, in contemporary physics, is not the inert and homogeneous substratum employed in both Newton's and Maxwell's theories; some of its new properties are not unlike those once attributed to the ether; we may someday come to know what an electric displacement is.

By shifting emphasis from the cognitive to the normative functions of paradigms, the preceding examples enlarge our understanding of the ways in which paradigms give form to the scientific life. Previously, we had principally examined the paradigm's role as a vehicle for scientific theory.

electric field, and so on. In this way, electromagnetic waves (such as radio waves and light rays) can propagate in a vacuum. But what is the medium, the so-called ether, in which these waves propagate? As Kuhn remarks, despite the empirical success of Maxwell's equations, all attempts to construct a consistent mechanical model of the ether met with failure. Eventually, with the successful integration of Maxwell's equations within the special theory of relativity, physicists stopped asking what the displacement current in free space is a displacement of and contented themselves with defining D operationally, in terms of quantities that can be observed and measured.

In that role it functions by telling the scientist about the entities that nature does and does not contain and about the ways in which those entities behave. That information provides a map whose details are elucidated by mature scientific research. And since nature is too complex and varied to be explored at random, that map is as essential as observation and experiment to science's continuing development. Through the theories they embody, paradigms prove to be constitutive of the research activity. They are also, however, constitutive of science in other respects, and that is now the point. In particular, our most recent examples show that paradigms provide scientists not only with a map but also with some of the directions essential for map-making. In learning a paradigm the scientist acquires theory, methods, and standards together, usually in an inextricable mixture. Therefore, when paradigms change, there are usually significant shifts in the criteria determining the legitimacy both of problems and of proposed solutions.

That observation returns us to the point from which this section began, for it provides our first explicit indication of why the choice between competing paradigms regularly raises questions that cannot be resolved by the criteria of normal science. To the extent, as significant as it is incomplete, that two scientific schools disagree about what is a problem and what a solution, they will inevitably talk through each other when debating the relative merits of their respective paradigms. In the partially circular arguments that regularly result, each paradigm will be shown to satisfy more or less the criteria that it dictates for itself and to fall short of a few of those dictated by its opponent. There are other reasons, too, for the incompleteness of logical contact that consistently characterizes paradigm debates. For example, since no paradigm ever solves all the problems it defines and since no two paradigms leave all the same problems unsolved, paradigm debates always involve the question: Which problems is it more significant to have solved? Like the issue of competing standards, that question of values can be answered only in terms of criteria that lie outside of normal science altogether, and it is that recourse to external criteria that most obviously makes paradigm debates revolutionary. . . .

■ | Notes

1. Silvanus P. Thompson, *Life of William Thomson Baron Kelvin of Largs* (London, 1910), I, 266–81.
2. See, for example, the remarks by P. P. Wiener in *Philosophy of Science*, XXV (1958), 298.
3. James B. Conant, *Overthrow of the Phlogiston Theory* (Cambridge, 1950), pp. 13–16; and J. R. Partington, *A Short History of Chemistry* (2d ed.; London, 1951), pp. 85–88. The fullest and most sympathetic account of the phlogiston

theory's achievements is by H. Metzger, *Newton, Stahl, Boerhaave et la doctrine chimique* (Paris, 1930), Part II.

4. Compare the conclusions reached through a very different sort of analysis by R. B. Braithwaite, *Scientific Explanation* (Cambridge, 1953), pp. 50-87. esp. p. 76.

5. For corpuscularism in general, see Marie Boas, "The Establishment of the Mechanical Philosophy," *Osiris*, X (1952), 412-541. For the effect of particle-shape on taste, see *ibid.*, p. 483.

6. R. Dugas, *La mécanique au XVII^e siècle* (Neuchâtel, 1954), pp. 177-85, 284-98, 345-56.

7. I. B. Cohen, *Franklin and Newton: An Inquiry into Speculative Newtonian Experimental Science and Franklin's Work in Electricity as an Example Thereof* (Philadelphia, 1956), chaps. vi-vii.

8. For electricity, see *ibid.*, chaps. viii-ix. For chemistry, see Metzger, *op. cit.*, Part I.

9. E. Meyerson, *Identity and Reality* (New York, 1930), chap. x.

10. E. T. Whittaker, *A History of the Theories of Aether and Electricity*, II (London, 1953), 28-30.

11. For a brilliant and entirely up-to-date attempt to fit scientific development into this Procrustean bed, see C. C. Gillispie, *The Edge of Objectivity: An Essay in the History of Scientific Ideas* (Princeton, 1960).

THOMAS S. KUHN

Objectivity, Value Judgment, and Theory Choice

In the penultimate chapter of a controversial book first published fifteen years ago, I considered the ways scientists are brought to abandon one time-honored theory or paradigm in favor of another. Such decision problems, I wrote, "cannot be resolved by proof." To discuss their mechanism is, therefore, to talk "about techniques of persuasion, or about argument and counterargument in a situation in which there can be no proof." Under these circumstances, I continued, "lifelong resistance [to a new theory] . . . is not a violation of scientific standards. . . . Though the historian can always find men—Priestley, for instance—who were unreasonable to resist for as long as they did, he will not find a point at which resistance becomes illogical or unscientific."¹ Statements of that sort obviously raise the question of why, in the absence of binding criteria for scientific choice, both the number of solved scientific problems and the precision of individual problem solutions should increase so markedly with the passage of time. Confronting that issue, I sketched in my closing chapter a number of characteristics that scientists share by virtue of the training which licenses their membership in one or another community of specialists. In the absence of criteria able to dictate the choice of each individual, I argued, we do well to trust the collective judgment of scientists trained in this way. "What better criterion could there be," I asked rhetorically, "than the decision of the scientific group?"²

A number of philosophers have greeted remarks like these in a way that continues to surprise me. My views, it is said, make of theory choice "a matter for mob psychology."³ Kuhn believes, I am told, that "the decision of a scientific group to adopt a new paradigm cannot be based on

FROM Thomas S. Kuhn, *The Essential Tension: Selected Studies in Scientific Tradition and Change* (Chicago: University of Chicago Press, 1977), 320–39. This essay was originally presented as the Machette Lecture, delivered at Furman University, Greenville, S.C., on 30 November 1973.

good reasons of any kind, factual or otherwise."⁴ The debates surrounding such choices must, my critics claim, be for me "mere persuasive displays without deliberative substance."⁵ Reports of this sort manifest total misunderstanding, and I have occasionally said as much in papers directed primarily to other ends. But those passing protestations have had negligible effect, and the misunderstandings continue to be important. I conclude that it is past time for me to describe, at greater length and with greater precision, what has been on my mind when I have uttered statements like the ones with which I just began. If I have been reluctant to do so in the past, that is largely because I have preferred to devote attention to areas in which my views diverge more sharply from those currently received than they do with respect to theory choice.

What, I ask to begin with, are the characteristics of a good scientific theory? Among a number of quite usual answers I select five, not because they are exhaustive, but because they are individually important and collectively sufficiently varied to indicate what is at stake. First, a theory should be accurate: within its domain, that is, consequences deducible from a theory should be in demonstrated agreement with the results of existing experiments and observations. Second, a theory should be consistent, not only internally or with itself, but also with other currently accepted theories applicable to related aspects of nature. Third, it should have broad scope: in particular, a theory's consequences should extend far beyond the particular observations, laws, or subtheories it was initially designed to explain. Fourth, and closely related, it should be simple, bringing order to phenomena that in its absence would be individually isolated and, as a set, confused. Fifth—a somewhat less standard item, but one of special importance to actual scientific decisions—a theory should be fruitful of new research findings: it should, that is, disclose new phenomena or previously unnoted relationships among those already known.⁶ These five characteristics—accuracy, consistency, scope, simplicity, and fruitfulness—are all standard criteria for evaluating the adequacy of a theory. If they had not been, I would have devoted far more space to them in my book, for I agree entirely with the traditional view that they play a vital role when scientists must choose between an established theory and an upstart competitor. Together with others of much the same sort, they provide *the* shared basis for theory choice.

Nevertheless, two sorts of difficulties are regularly encountered by the men who must use these criteria in choosing, say, between Ptolemy's astronomical theory and Copernicus's, between the oxygen and phlogiston theories of combustion, or between Newtonian mechanics and the quantum theory. Individually the criteria are imprecise: individuals may legitimately differ about their application to concrete cases. In addition, when deployed together, they repeatedly prove to conflict with one another; accuracy may, for example, dictate the choice of one theory, scope the

choice of its competitor. Since these difficulties, especially the first, are also relatively familiar, I shall devote little time to their elaboration. Though my argument does demand that I illustrate them briefly, my views will begin to depart from those long current only after I have done so.

Begin with accuracy, which for present purposes I take to include not only quantitative agreement but qualitative as well. Ultimately it proves the most nearly decisive of all the criteria, partly because it is less equivocal than the others but especially because predictive and explanatory powers, which depend on it, are characteristics that scientists are particularly unwilling to give up. Unfortunately, however, theories cannot always be discriminated in terms of accuracy. Copernicus's system, for example, was not more accurate than Ptolemy's until drastically revised by Kepler more than sixty years after Copernicus's death. If Kepler or someone else had not found other reasons to choose heliocentric astronomy, those improvements in accuracy would never have been made, and Copernicus's work might have been forgotten. More typically, of course, accuracy does permit discriminations, but not the sort that lead regularly to unequivocal choice. The oxygen theory, for example, was universally acknowledged to account for observed weight relations in chemical reactions, something the phlogiston theory had previously scarcely attempted to do. But the phlogiston theory, unlike its rival, could account for the metals' being much more alike than the ores from which they were formed. One theory thus matched experience better in one area, the other in another.* To choose between them on the basis of accuracy, a scientist would need to decide the area in which accuracy was more significant. About that matter chemists could and did differ without violating any of the criteria outlined above, or any others yet to be suggested.

However important it may be, therefore, accuracy by itself is seldom or never a sufficient criterion for theory choice. Other criteria must func-

* Supporters of the phlogiston theory argued that metals are similar because they all contain phlogiston, which is released as heat and fire when metals burn in air. (This argument is not very impressive, since it fails to explain why carbon and other nonmetallic combustible materials, which are also supposed to be rich in phlogiston, are not at all like metals.) Lavoisier's oxygen theory offered no explanation of why metals resemble each other, but it did predict that all metals become heavier when they burn, and this prediction was confirmed by weighing experiments. Metals gain weight when they burn because burning involves the chemical combination of the metal with the oxygen in the air to form an oxide. The heat associated with oxidation was attributed by Lavoisier to the release of caloric fluid that, according to his theory, surrounds the particles of oxygen. This is a typical example of what has come to be known as a "Kuhn loss": when one paradigm replaces another, not every problem that was solved by the old paradigm can be solved by the new one, even though the new paradigm solves problems that the old one either ignored or could not solve. Modern science has finally succeeded in explaining the similarity among metals: metals are shiny, conduct electricity, etc. because they all contain free electrons that are not bound to individual atoms.

tion as well, but they do not eliminate problems. To illustrate I select just two—consistency and simplicity—asking how they functioned in the choice between the heliocentric and geocentric systems. As astronomical theories both Ptolemy's and Copernicus's were internally consistent, but their relation to related theories in other fields was very different. The stationary central earth was an essential ingredient of received physical theory, a tight-knit body of doctrine which explained, among other things, how stones fall, how water pumps function, and why the clouds move slowly across the skies. Heliocentric astronomy, which required the earth's motion, was inconsistent with the existing scientific explanation of these and other terrestrial phenomena. The consistency criterion, by itself, therefore, spoke unequivocally for the geocentric tradition.

Simplicity, however, favored Copernicus, but only when evaluated in a quite special way. If, on the one hand, the two systems were compared in terms of the actual computational labor required to predict the position of a planet at a particular time, then they proved substantially equivalent. Such computations were what astronomers did, and Copernicus's system offered them no labor-saving techniques; in that sense it was not simpler than Ptolemy's. If, on the other hand, one asked about the amount of mathematical apparatus required to explain, not the detailed quantitative motions of the planets, but merely their gross qualitative features—limited elongation, retrograde motion, and the like—then, as every schoolchild knows, Copernicus required only one circle per planet, Ptolemy two. In that sense the Copernican theory was the simpler, a fact vitally important to the choices made by both Kepler and Galileo and thus essential to the ultimate triumph of Copernicanism. But that sense of simplicity was not the only one available, nor even the one most natural to professional astronomers, men whose task was the actual computation of planetary position.

Because time is short and I have multiplied examples elsewhere, I shall here simply assert that these difficulties in applying standard criteria of choice are typical and that they arise no less forcefully in twentieth-century situations than in the earlier and better-known examples I have just sketched. When scientists must choose between competing theories, two men fully committed to the same list of criteria for choice may nevertheless reach different conclusions. Perhaps they interpret simplicity differently or have different convictions about the range of fields within which the consistency criterion must be met. Or perhaps they agree about these matters but differ about the relative weights to be accorded to these or to other criteria when several are deployed together. With respect to divergences of this sort, no set of choice criteria yet proposed is of any use. One can explain, as the historian characteristically does, why particular men made particular choices at particular times. But for that purpose one must go beyond the list of shared criteria to characteristics of the individuals who make the choice. One must, that is, deal with character-

istics which vary from one scientist to another without thereby in the least jeopardizing their adherence to the canons that make science scientific. Though such canons do exist and should be discoverable (doubtless the criteria of choice with which I began are among them), they are not by themselves sufficient to determine the decisions of individual scientists. For that purpose the shared canons must be fleshed out in ways that differ from one individual to another.

Some of the differences I have in mind result from the individual's previous experience as a scientist. In what part of the field was he at work when confronted by the need to choose? How long had he worked there; how successful had he been; and how much of his work depended on concepts and techniques challenged by the new theory? Other factors relevant to choice lie outside the sciences. Kepler's early election of Copernicanism was due in part to his immersion in the Neoplatonic and Hermetic movements of his day; German Romanticism predisposed those it affected toward both recognition and acceptance of energy conservation; nineteenth-century British social thought had a similar influence on the availability and acceptability of Darwin's concept of the struggle for existence. Still other significant differences are functions of personality. Some scientists place more premium than others on originality and are correspondingly more willing to take risks; some scientists prefer comprehensive, unified theories to precise and detailed problem solutions of apparently narrower scope. Differentiating factors like these are described by my critics as subjective and are contrasted with the shared or objective criteria from which I began. Though I shall later question that use of terms, let me for the moment accept it. My point is, then, that every individual choice between competing theories depends on a mixture of objective and subjective factors, or of shared and individual criteria. Since the latter have not ordinarily figured in the philosophy of science, my emphasis upon them has made my belief in the former hard for my critics to see.

What I have said so far is primarily simply descriptive of what goes on in the sciences at times of theory choice. As description, furthermore, it has not been challenged by my critics, who reject instead my claim that these facts of scientific life have philosophic import. Taking up that issue, I shall begin to isolate some, though I think not vast, differences of opinion. Let me begin by asking how philosophers of science can for so long have neglected the subjective elements which, they freely grant, enter regularly into the actual theory choices made by individual scientists? Why have these elements seemed to them an index only of human weakness, not at all of the nature of scientific knowledge?

One answer to that question is, of course, that few philosophers, if any, have claimed to possess either a complete or an entirely well-articulated list of criteria. For some time, therefore, they could reasonably expect

that further research would eliminate residual imperfections and produce an algorithm able to dictate rational, unanimous choice. Pending that achievement, scientists would have no alternative but to supply subjectively what the best current list of objective criteria still lacked. That some of them might still do so even with a perfected list at hand would then be an index only of the inevitable imperfection of human nature.

That sort of answer may still prove to be correct, but I think no philosopher still expects that it will. The search for algorithmic decision procedures has continued for some time and produced both powerful and illuminating results. But those results all presuppose that individual criteria of choice can be unambiguously stated and also that, if more than one proves relevant, an appropriate weight function is at hand for their joint application. Unfortunately, where the choice at issue is between scientific theories, little progress has been made toward the first of these desiderata and none toward the second. Most philosophers of science would, therefore, I think, now regard the sort of algorithm which has traditionally been sought as a not quite attainable ideal. I entirely agree and shall henceforth take that much for granted.

Even an ideal, however, if it is to remain credible, requires some demonstrated relevance to the situations in which it is supposed to apply. Claiming that such demonstration requires no recourse to subjective factors, my critics seem to appeal, implicitly or explicitly, to the well-known distinction between the contexts of discovery and of justification.^{7*} They concede, that is, that the subjective factors I invoke play a significant role in the discovery or invention of new theories, but they also insist that that inevitably intuitive process lies outside of the bounds of philosophy of science and is irrelevant to the question of scientific objectivity. Objectivity enters science, they continue, through the processes by which theories are tested, justified, or judged. Those processes do not, or at least need not, involve subjective factors at all. They can be governed by a set of (objective) criteria shared by the entire group competent to judge.

I have already argued that that position does not fit observations of scientific life and shall now assume that that much has been conceded. What is now at issue is a different point: whether or not this invocation of the distinction between contexts of discovery and of justification provides even a plausible and useful idealization. I think it does not and can best make my point by suggesting first a likely source of its apparent cogency. I suspect that my critics have been misled by science pedagogy or

* The "well-known" distinction to which Kuhn refers—between the contexts of discovery and justification—has been endorsed by many philosophers of science. The phrases *context of discovery* and *context of justification* were coined by Hans Reichenbach in *Experience and Prediction* (Chicago: University of Chicago Press, 1938), ch. 1. For further discussion, see "The Problem of Description," in the commentary on chapter 4 below.

what I have elsewhere called textbook science. In science teaching, theories are presented together with exemplary applications, and those applications may be viewed as evidence. But that is not their primary pedagogic function (science students are distressingly willing to receive the word from professors and texts). Doubtless *some* of them were *part* of the evidence at the time actual decisions were being made, but they represent only a fraction of the considerations relevant to the decision process. The context of pedagogy differs almost as much from the context of justification as it does from that of discovery.

Full documentation of that point would require longer argument than is appropriate here, but two aspects of the way in which philosophers ordinarily demonstrate the relevance of choice criteria are worth noting. Like the science textbooks on which they are often modelled, books and articles on the philosophy of science refer again and again to the famous crucial experiments:^{*} Foucault's pendulum, which demonstrates the motion of the earth; Cavendish's demonstration of gravitational attraction; or Fizeau's measurement of the relative speed of sound in water and air. These experiments are *paradigms* of good reason for scientific choice; they illustrate the most effective of all the sorts of argument which could be available to a scientist uncertain which of two theories to follow; they are vehicles for the transmission of criteria of choice. But they also have another characteristic in common. By the time they were performed no scientist still needed to be convinced of the validity of the theory their outcome is now used to demonstrate. Those decisions had long since been made on the basis of significantly more equivocal evidence. The exemplary crucial experiments to which philosophers again and again refer would have been historically relevant to theory choice only if they had yielded unexpected results. Their use as illustrations provides needed economy to science pedagogy, but they scarcely illuminate the character of the choices that scientists are called upon to make.

Standard philosophical illustrations of scientific choice have another troublesome characteristic. The only arguments discussed are, as I have previously indicated, the ones favorable to the theory that, in fact, ultimately triumphed. Oxygen, we read, could explain weight relations, phlogiston could not; but nothing is said about the phlogiston theory's power or about the oxygen theory's limitations. Comparisons of Ptolemy's theory

^{*} A crucial experiment is one that conclusively falsifies one of two rival theories or hypotheses, thereby establishing its rival as well confirmed or true. Thus, for example, Kuhn describes Foucault's pendulum as a crucial experiment because it conclusively refutes the hypothesis that the earth is stationary, thereby "demonstrating" the motion of the earth. For further discussion, see the section, "Why Crucial Experiments Are Impossible in Physics," in the commentary on chapter 3.

with Copernicus's proceed in the same way. Perhaps these examples should not be given since they contrast a developed theory with one still in its infancy. But philosophers regularly use them nonetheless. If the only result of their doing so were to simplify the decision situation, one could not object. Even historians do not claim to deal with the full factual complexity of the situations they describe. But these simplifications emasculate by making choice totally unproblematic. They eliminate, that is, one essential element of the decision situations that scientists must resolve if their field is to move ahead. In those situations there are always at least some good reasons for each possible choice. Considerations relevant to the context of discovery are then relevant to justification as well; scientists who share the concerns and sensibilities of the individual who discovers a new theory are ipso facto likely to appear disproportionately frequently among that theory's first supporters. That is why it has been difficult to construct algorithms for theory choice, and also why such difficulties have seemed so thoroughly worth resolving. Choices that present problems are the ones philosophers of science need to understand. Philosophically interesting decision procedures must function where, in their absence, the decision might still be in doubt.

That much I have said before, if only briefly. Recently, however, I have recognized another, subtler source for the apparent plausibility of my critics' position. To present it, I shall briefly describe a hypothetical dialogue with one of them. Both of us agree that each scientist chooses between competing theories by deploying some Bayesian algorithm which permits him to compute a value for $p(T,E)$, i.e., for the probability of a theory T on the evidence E available both to him and to the other members of his professional group at a particular period of time. "Evidence," furthermore, we both interpret broadly to include such considerations as simplicity and fruitfulness. My critic asserts, however, that there is only one such value of p , that corresponding to objective choice, and he believes that all rational members of the group must arrive at it. I assert, on the other hand, for reasons previously given, that the factors he calls objective are insufficient to determine in full any algorithm at all. For the sake of the discussion I have conceded that each individual has an algorithm and that all their algorithms have much in common. Nevertheless, I continue to hold that the algorithms of individuals are all ultimately different by virtue of the subjective considerations with which each must complete the objective criteria before any computations can be done. If my hypothetical critic is liberal, he may now grant that these subjective differences do play a role in determining the hypothetical algorithm on which each individual relies during the early stages of the competition between rival theories. But he is also likely to claim that, as evidence increases with the passage of time, the algorithms of different individuals converge to the algorithm of objective choice with which his presentation

began. For him the increasing unanimity of individual choices is evidence for their increasing objectivity and thus for the elimination of subjective elements from the decision process.

So much for the dialogue, which I have, of course, contrived to disclose the non sequitur underlying an apparently plausible position. What converges as the evidence changes over time need only be the values of p that individuals compute from their individual algorithms. Conceivably those algorithms themselves also become more alike with time, but the ultimate unanimity of theory choice provides no evidence whatsoever that they do so. If subjective factors are required to account for the decisions that initially divide the profession, they may still be present later when the profession agrees. Though I shall not here argue the point, consideration of the occasions on which a scientific community divides suggests that they actually do so.

My argument has so far been directed to two points. It first provided evidence that the choices scientists make between competing theories depend not only on shared criteria—those my critics call objective—but also on idiosyncratic factors dependent on individual biography and personality. The latter are, in my critics' vocabulary, subjective, and the second part of my argument has attempted to bar some likely ways of denying their philosophic import. Let me now shift to a more positive approach, returning briefly to the list of shared criteria—accuracy, simplicity, and the like—with which I began. The considerable effectiveness of such criteria does not, I now wish to suggest, depend on their being sufficiently articulated to dictate the choice of each individual who subscribes to them. Indeed, if they were articulated to that extent, a behavior mechanism fundamental to scientific advance would cease to function. What the tradition sees as eliminable imperfections in its rules of choice I take to be in part responses to the essential nature of science.

As so often, I begin with the obvious. Criteria that influence decisions without specifying what those decisions must be are familiar in many aspects of human life. Ordinarily, however, they are called, not criteria or rules, but maxims, norms, or values. Consider maxims first. The individual who invokes them when choice is urgent usually finds them frustratingly vague and often also in conflict one with another. Contrast "He who hesitates is lost" with "Look before you leap," or compare "Many hands make light work" with "Too many cooks spoil the broth." Individually maxims dictate different choices, collectively none at all. Yet no one suggests that supplying children with contradictory tags like these is irrelevant to their education. Opposing maxims alter the nature of the decision to be made, highlight the essential issues it presents, and point to those remaining aspects of the decision for which each individual must take responsibility himself. Once invoked, maxims like these alter the nature of the decision process and can thus change its outcome.

Values and norms provide even clearer examples of effective guidance in the presence of conflict and equivocation. Improving the quality of life is a value, and a car in every garage once followed from it as a norm. But quality of life has other aspects, and the old norm has become problematic. Or again, freedom of speech is a value, but so is preservation of life and property. In application, the two often conflict, so that judicial soul-searching, which still continues, has been required to prohibit such behavior as inciting to riot or shouting fire in a crowded theater. Difficulties like these are an appropriate source for frustration, but they rarely result in charges that values have no function or in calls for their abandonment. That response is barred to most of us by an acute consciousness that there are societies with other values and that these value differences result in other ways of life, other decisions about what may and what may not be done.

I am suggesting, of course, that the criteria of choice with which I began function not as rules, which determine choice, but as values, which influence it. Two men deeply committed to the same values may nevertheless, in particular situations, make different choices as, in fact, they do. But that difference in outcome ought not to suggest that the values scientists share are less than critically important either to their decisions or to the development of the enterprise in which they participate. Values like accuracy, consistency, and scope may prove ambiguous in application, both individually and collectively; they may, that is, be an insufficient basis for a shared algorithm of choice. But they do specify a great deal: what each scientist must consider in reaching a decision, what he may and may not consider relevant, and what he can legitimately be required to report as the basis for the choice he has made. Change the list, for example by adding social utility as a criterion, and some particular choices will be different, more like those one expects from an engineer. Subtract accuracy of fit to nature from the list, and the enterprise that results may not resemble science at all, but perhaps philosophy instead. Different creative disciplines are characterized, among other things, by different sets of shared values. If philosophy and engineering lie too close to the sciences, think of literature or the plastic arts. Milton's failure to set *Paradise Lost* in a Copernican universe does not indicate that he agreed with Ptolemy but that he had things other than science to do.

Recognizing that criteria of choice can function as values when incomplete as rules has, I think, a number of striking advantages. First, as I have already argued at length, it accounts in detail for aspects of scientific behavior which the tradition has seen as anomalous or even irrational. More important, it allows the standard criteria to function fully in the earliest stages of theory choice, the period when they are most needed but when, on the traditional view, they function badly or not at all. Copernicus was responding to them during the years required to convert heliocentric astronomy from a global conceptual scheme to mathematical machinery

for predicting planetary position. Such predictions were what astronomers valued; in their absence, Copernicus would scarcely have been heard, something which had happened to the idea of a moving earth before. That his own version convinced very few is less important than his acknowledgment of the basis on which judgments would have to be reached if heliocentricism were to survive. Though idiosyncrasy must be invoked to explain why Kepler and Galileo were early converts to Copernicus's system, the gaps filled by their efforts to perfect it were specified by shared values alone.

That point has a corollary which may be more important still. Most newly suggested theories do not survive. Usually the difficulties that evoked them are accounted for by more traditional means. Even when this does not occur, much work, both theoretical and experimental, is ordinarily required before the new theory can display sufficient accuracy and scope to generate widespread conviction. In short, before the group accepts it, a new theory has been tested over time by the research of a number of men, some working within it, others within its traditional rival. Such a mode of development, however, *requires* a decision process which permits rational men to disagree, and such disagreement would be barred by the shared algorithm which philosophers have generally sought. If it were at hand, all conforming scientists would make the same decision at the same time. With standards for acceptance set too low, they would move from one attractive global viewpoint to another, never giving traditional theory an opportunity to supply equivalent attractions. With standards set higher, no one satisfying the criterion of rationality would be inclined to try out the new theory, to articulate it in ways which showed its fruitfulness or displayed its accuracy and scope. I doubt that science would survive the change. What from one viewpoint may seem the looseness and imperfection of choice criteria conceived as rules may, when the same criteria are seen as values, appear an indispensable means of spreading the risk which the introduction or support of novelty always entails.

Even those who have followed me this far will want to know how a value-based enterprise of the sort I have described can develop as a science does, repeatedly producing powerful new techniques for prediction and control. To that question, unfortunately, I have no answer at all, but that is only another way of saying that I make no claim to have solved the problem of induction. If science did progress by virtue of some shared and binding algorithm of choice, I would be equally at a loss to explain its success. The lacuna is one I feel acutely, but its presence does not differentiate my position from the tradition.

It is, after all, no accident that my list of the values guiding scientific choice is, as nearly as makes any difference, identical with the tradition's list of rules dictating choice. Given any concrete situation to which the philosopher's rules could be applied, my values would function like his rules, producing the same choice. Any justification of induction, any ex-

planation of why the rules worked, would apply equally to my values. Now consider a situation in which choice by shared rules proves impossible, not because the rules are wrong but because they are, as rules, intrinsically incomplete. Individuals must then still choose and be guided by the rules (now values) when they do so. For that purpose, however, each must first flesh out the rules, and each will do so in a somewhat different way even though the decision dictated by the variously completed rules may prove unanimous. If I now assume, in addition, that the group is large enough so that individual differences distribute on some normal curve, then any argument that justifies the philosopher's choice by rule should be immediately adaptable to my choice by value. A group too small, or a distribution excessively skewed by external historical pressures, would, of course, prevent the argument's transfer.⁶ But those are just the circumstances under which scientific progress is itself problematic. The transfer is not then to be expected.

I shall be glad if these references to a normal distribution of individual differences and to the problem of induction make my position appear very close to more traditional views. With respect to theory choice, I have never thought my departures large and have been correspondingly startled by such charges as "mob psychology," quoted at the start. It is worth noting, however, that the positions are not quite identical, and for that purpose an analogy may be helpful. Many properties of liquids and gases can be accounted for on the kinetic theory by supposing that all molecules travel at the same speed. Among such properties are the regularities known as Boyle's and Charles's law. Other characteristics, most obviously evaporation, cannot be explained in so simple a way. To deal with them one must assume that molecular speeds differ, that they are distributed at random, governed by the laws of chance. What I have been suggesting here is that theory choice, too, can be explained only in part by a theory which attributes the same properties to all the scientists who must do the choosing. Essential aspects of the process generally known as verification will be understood only by recourse to the features with respect to which men may differ while still remaining scientists. The tradition takes it for granted that such features are vital to the process of discovery, which it at once and for that reason rules out of philosophical bounds. That they may have significant functions also in the philosophically central problem of justifying theory choice is what philosophers of science have to date categorically denied.

What remains to be said can be grouped in a somewhat miscellaneous epilogue. For the sake of clarity and to avoid writing a book, I have throughout this paper utilized some traditional concepts and locutions about the viability of which I have elsewhere expressed serious doubts. For those who know the work in which I have done so, I close by indicating three aspects of what I have said which would better represent my views

if cast in other terms, simultaneously indicating the main directions in which such recasting should proceed. The areas I have in mind are: value invariance, subjectivity, and partial communication. If my views of scientific development are novel—a matter about which there is legitimate room for doubt—it is in areas such as these, rather than theory choice, that my main departures from tradition should be sought.

Throughout this paper I have implicitly assumed that, whatever their initial source, the criteria or values deployed in theory choice are fixed once and for all, unaffected by their participation in transitions from one theory to another. Roughly speaking, but only very roughly, I take that to be the case. If the list of relevant values is kept short (I have mentioned five, not all independent) and if their specification is left vague, then such values as accuracy, scope, and fruitfulness are permanent attributes of science. But little knowledge of history is required to suggest that both the application of these values and, more obviously, the relative weights attached to them have varied markedly with time and also with the field of application. Furthermore, many of these variations in value have been associated with particular changes in scientific theory. Though the experience of scientists provides no philosophical justification for the values they deploy (such justification would solve the problem of induction), those values are in part learned from that experience, and they evolve with it.

The whole subject needs more study (historians have usually taken scientific values, though not scientific methods, for granted), but a few remarks will illustrate the sort of variations I have in mind. Accuracy, as a value, has with time increasingly denoted quantitative or numerical agreement, sometimes at the expense of qualitative. Before early modern times, however, accuracy in that sense was a criterion only for astronomy, the science of the celestial region. Elsewhere it was neither expected nor sought. During the seventeenth century, however, the criterion of numerical agreement was extended to mechanics, during the late eighteenth and early nineteenth centuries to chemistry and such other subjects as electricity and heat, and in this century to many parts of biology. Or think of utility, an item of value not on my initial list. It too has figured significantly in scientific development, but far more strongly and steadily for chemists than for, say, mathematicians and physicists. Or consider scope. It is still an important scientific value, but important scientific advances have repeatedly been achieved at its expense, and the weight attributed to it at times of choice has diminished correspondingly.

What may seem particularly troublesome about changes like these is, of course, that they ordinarily occur in the aftermath of a theory change. One of the objections to Lavoisier's new chemistry was the roadblocks with which it confronted the achievement of what had previously been one of chemistry's traditional goals: the explanation of qualities, such as color and texture, as well as of their changes. With the acceptance of

Lavoisier's theory such explanations ceased for some time to be a value for chemists; the ability to explain qualitative variation was no longer a criterion relevant to the evaluation of chemical theory. Clearly, if such value changes had occurred as rapidly or been as complete as the theory changes to which they related, then theory choice would be value choice, and neither could provide justification for the other. But, historically, value change is ordinarily a belated and largely unconscious concomitant of theory choice, and the former's magnitude is regularly smaller than the latter's. For the functions I have here ascribed to values, such relative stability provides a sufficient basis. The existence of a feedback loop through which theory change affects the values which led to that change does not make the decision process circular in any damaging sense.

About a second respect in which my resort to tradition may be misleading, I must be far more tentative. It demands the skills of an ordinary language philosopher, which I do not possess. Still, no very acute ear for language is required to generate discomfort with the ways in which the terms "objectivity" and, more especially, "subjectivity" have functioned in this paper. Let me briefly suggest the respects in which I believe language has gone astray. "Subjective" is a term with several established uses: in one of these it is opposed to "objective," in another to "judgmental." When my critics describe the idiosyncratic features to which I appeal as subjective, they resort, erroneously I think, to the second of these senses. When they complain that I deprive science of objectivity, they conflate that second sense of subjective with the first.

A standard application of the term "subjective" is to matters of taste, and my critics appear to suppose that that is what I have made of theory choice. But they are missing a distinction standard since Kant when they do so. Like sensation reports, which are also subjective in the sense now at issue, matters of taste are undiscussable. Suppose that, leaving a movie theater with a friend after seeing a western, I exclaim: "How I liked that terrible potboiler!" My friend, if he disliked the film, may tell me I have low tastes, a matter about which, in these circumstances, I would readily agree. But, short of saying that I lied, he cannot disagree with my report that I liked the film or try to persuade me that what I said about my reaction was wrong. What is discussable in my remark is not my characterization of my internal state, my exemplification of taste, but rather my *judgment* that the film was a potboiler. Should my friend disagree on that point, we may argue most of the night, each comparing the film with good or great ones we have seen, each revealing, implicitly or explicitly, something about how he *judges* cinematic merit, about his aesthetic. Though one of us may, before retiring, have persuaded the other, he need not have done so to demonstrate that our difference is one of judgment, not taste.

Evaluations or choices of theory have, I think, exactly this character. Not that scientists never say merely, I like such and such a theory, or I do not. After 1926 Einstein said little more than that about his opposition to

the quantum theory.* But scientists may always be asked to explain their choices, to exhibit the bases for their judgments. Such judgments are eminently discussable, and the man who refuses to discuss his own cannot expect to be taken seriously. Though there are, very occasionally, leaders of scientific taste, their existence tends to prove the rule. Einstein was one of the few, and his increasing isolation from the scientific community in later life shows how very limited a role taste alone can play in theory choice. Bohr, unlike Einstein, did discuss the bases for his judgment, and he carried the day. If my critics introduce the term "subjective" in a sense that opposes it to judgmental—thus suggesting that I make theory choice undiscussable, a matter of taste—they have seriously mistaken my position.

Turn now to the sense in which "subjectivity" is opposed to "objectivity," and note first that it raises issues quite separate from those just discussed. Whether my taste is low or refined, my report that I liked the film is objective unless I have lied. To my judgment that the film was a potboiler, however, the objective-subjective distinction does not apply at all, at least not obviously and directly. When my critics say I deprive theory choice of objectivity, they must, therefore, have recourse to some very different sense of subjective, presumably the one in which bias and personal likes or dislikes function instead of, or in the face of, the actual facts. But that sense of subjective does not fit the process I have been describing any better than the first. Where factors dependent on individual biography or personality must be introduced to make values applicable, no standards of factuality or actuality are being set aside. Conceivably my discussion of theory choice indicates some limitations of objectivity, but not by isolating elements properly called subjective. Nor am I even quite content with the notion that what I have been displaying are limitations. Objectivity ought to be analyzable in terms of criteria like accuracy and consistency. If these criteria do not supply all the guidance that we have customarily expected of them, then it may be the meaning rather than the limits of objectivity that my argument shows.

Turn, in conclusion, to a third respect, or set of respects, in which this paper needs to be recast. I have assumed throughout that the discussions surrounding theory choice are unproblematic, that the facts appealed

* Presumably Kuhn meant "1936," given that from 1927 to 1936 Einstein and Bohr carried on a debate about quantum mechanics that has been described as "one of the great intellectual disputes in the history of science." The debate culminated in the famous EPR paper of 1936 in which the authors (Einstein, Podolsky, and Rosen) argued that quantum mechanics could not give a complete description of reality. During this period, Einstein was an articulate and relentless critic of the so-called Copenhagen interpretation of quantum mechanics. See Arthur Fine, *The Shaky Game: Einstein, Realism, and the Quantum Theory* (Chicago, Ill.: University of Chicago Press, 1986), and Dugald Murdoch, *Niels Bohr's Philosophy of Physics* (Cambridge: University of Cambridge Press, 1987). The quotation is from page 155 of Murdoch's book.

to in such discussions are independent of theory, and that the discussions' outcome is appropriately called a choice. Elsewhere I have challenged all three of these assumptions, arguing that communication between proponents of different theories is inevitably partial, that what each takes to be facts depends in part on the theory he espouses, and that an individual's transfer of allegiance from theory to theory is often better described as conversion than as choice. Though all these theses are problematic as well as controversial, my commitment to them is undiminished. I shall not now defend them, but must at least attempt to indicate how what I have said here can be adjusted to conform with these more central aspects of my view of scientific development.

For that purpose I resort to an analogy I have developed in other places. Proponents of different theories are, I have claimed, like native speakers of different languages. Communication between them goes on by translation, and it raises all translation's familiar difficulties. That analogy is, of course, incomplete, for the vocabulary of the two theories may be identical, and most words function in the same ways in both. But some words in the basic as well as in the theoretical vocabularies of the two theories—words like "star" and "planet," "mixture" and "compound," or "force" and "matter"—do function differently. Those differences are unexpected and will be discovered and localized, if at all, only by repeated experience of communication breakdown. Without pursuing the matter further, I simply assert the existence of significant limits to what the proponents of different theories can communicate to one another. The same limits make it difficult or, more likely, impossible for an individual to hold both theories in mind together and compare them point by point with each other and with nature. That sort of comparison is, however, the process on which the appropriateness of any word like "choice" depends.

Nevertheless, despite the incompleteness of their communication, proponents of different theories can exhibit to each other, not always easily, the concrete technical results achievable by those who practice within each theory. Little or no translation is required to apply at least some value criteria to those results. (Accuracy and fruitfulness are most immediately applicable, perhaps followed by scope. Consistency and simplicity are far more problematic.) However incomprehensible the new theory may be to the proponents of tradition, the exhibit of impressive concrete results will persuade at least a few of them that they must discover how such results are achieved. For that purpose they must learn to translate, perhaps by treating already published papers as a Rosetta stone or, often more effective, by visiting the innovator, talking with him, watching him and his students at work. Those exposures may not result in the adoption of the theory; some advocates of the tradition may return home and attempt to adjust the old theory to produce equivalent results. But others, if the new theory is to survive, will find that at some point in the language-learning process they have ceased to translate and begun instead to speak the lan-

guage like a native. No process quite like choice has occurred, but they are practicing the new theory nonetheless. Furthermore, the factors that have led them to risk the conversion they have undergone are just the ones this paper has underscored in discussing a somewhat different process, one which, following the philosophical tradition, it has labelled theory choice.

Notes

1. *The Structure of Scientific Revolutions*, 2d ed. (Chicago, 1970), pp. 148, 151–52, 159. All the passages from which these fragments are taken appeared in the same form in the first edition, published in 1962.
2. *Ibid.*, p. 170.
3. Imre Lakatos, "Falsification and the Methodology of Scientific Research Programmes," in I. Lakatos and A. Musgrave, eds., *Criticism and the Growth of Knowledge* (Cambridge, 1970), pp. 91–195. The quoted phrase, which appears on p. 178, is italicized in the original.
4. Dudley Shapere, "Meaning and Scientific Change," in R. G. Colodny, ed., *Mind and Cosmos: Essays in Contemporary Science and Philosophy*, University of Pittsburgh Series in the Philosophy of Science, vol. 3 (Pittsburgh, 1966), pp. 41–85. The quotation will be found on p. 67.
5. Israel Scheffler, *Science and Subjectivity* (Indianapolis, 1967), p. 81.
6. The last criterion, fruitfulness, deserves more emphasis than it has yet received. A scientist choosing between two theories ordinarily knows that his decision will have a bearing on his subsequent research career. Of course he is especially attracted by a theory that promises the concrete successes for which scientists are ordinarily rewarded.
7. The least equivocal example of this position is probably the one developed in Scheffler, *Science and Subjectivity*, chap. 4.
8. If the group is small, it is more likely that random fluctuations will result in its members' sharing an atypical set of values and therefore making choices different from those that would be made by a larger and more representative group. External environment—intellectual, ideological, or economic—must systematically affect the value system of much larger groups, and the consequences can include difficulties in introducing the scientific enterprise to societies with inimical values or perhaps even the end of that enterprise within societies where it had once flourished. In this area, however, great caution is required. Changes in the environment where science is practiced can also have fruitful effects on research. Historians often resort, for example, to differences between national environments to explain why particular innovations were initiated and at first disproportionately pursued in particular countries, e.g., Darwinism in Britain, energy conservation in Germany. At present we know substantially nothing about the minimum requisites of the social milieu within which a sciencelike enterprise might flourish.

Rationality and Paradigm Change in Science

As we look back at the first responses of philosophers of science to Thomas Kuhn's classic *The Structure of Scientific Revolutions* [SSR], we are struck by their near unanimity toward the challenge that the book posed to the rationality of science. Kuhn's account of the paradigm changes that for him constituted scientific revolutions was taken by many to undermine the rationality of the scientific process itself. The metaphors of conversion and gestalt switch, the insistence that defenders of rival paradigms must inevitably fail to make contact with each other's viewpoints, struck those philosophical readers whose expectations were formed by later logical empiricism as a deliberate rejection of the basic requirements of effective reason giving in the natural sciences.

Kuhn responded to this reading of SSR in a lengthy Postscript to the second edition of his book in 1970 and in the reflective essay "Objectivity, Value Judgment, and Theory Choice" in 1977.¹ He labored to show that the implications of his new account of scientific change for the *rationality* of that change were far less radical than his critics were taking them to be. But his disavowals were not, in the main, taken as seriously as he had hoped they would be; the echoes of the rhetoric of SSR still lingered in people's minds. It seems worth returning to this ground, familiar though it may seem, in order to assess just what Kuhn ~~did~~ have to say about how paradigm change comes about in science. We will see that the radical thrust of his account of science was indeed not directed so much against the rationality of theory choice as against the epistemic, or truthlike, character of the theories so chosen.

FROM Paul Horwich, ed., *World Changes: Thomas Kuhn and the Nature of Science* (Cambridge, Mass.: MIT Press, 1993), 55–78.

1 | Good Reasons for Paradigm Change

The theme that recurs in Kuhn's discussions of paradigm change is a two-sided one. On one hand, he wanted to emphasize the fundamental role played by "good reasons" in motivating theory change in science. Notable among these is the perception of anomaly, the growing awareness that something is wrong, which makes it possible for alternatives to be seriously viewed as alternatives. On the other hand, these reasons are never coercive in their own right in forcing change; the reasons in favor of a new paradigm cannot *compel* assent. There is no precise point at which resistance to the change of paradigm becomes illogical.² Proponents of the new paradigm and defenders of the old one may each be able to lay claim to be acting "rationally"; the fact that neither side can persuade the other does not undermine the claim each can make to have good reasons for what they assert. "The point I have been trying to make," Kuhn says in the Postscript to *SSR*, "is a simple one, long familiar in philosophy of science. Debates over theory-choice cannot be cast in a form that fully resembles logical or mathematical proof. . . . Nothing about that relatively familiar thesis implies either that there are no good reasons for being persuaded or that those reasons are not ultimately decisive for the group. Nor does it even imply that the reasons for choice are different from those usually listed by philosophers of science: accuracy, simplicity, fruitfulness, and the like. What it should suggest, however, is that such reasons function as values and that they can thus be differently applied, individually and collectively, by men who concur in honoring them."³

It is with the implications of this thesis that I will be mainly concerned in this essay. The values a good theory is expected to embody enable comparisons to be made, even when the rival theories are incommensurable. Kuhn makes it clear that "incommensurable" for him does not imply "incomparable." *SSR*, he notes, "includes many explicit examples of comparisons between successive theories. I have never doubted either that they were possible or that they were essential at times of theory choice."⁴ What he wanted to emphasize, he says, is that "successive theories are incommensurable (which is not the same as incomparable) in the sense that the referents of some of the terms which occur in both are a function of the theory within which those terms appear," and hence that there is no neutral language available for purposes of comparison. Nonetheless, translation is in principle possible.⁵ But to translate another's theory is still not to make it one's own. "For that one must go native, discover that one is thinking and working in, not simply translating out of, a language that was previously foreign."⁶ And that transition cannot simply be willed, he maintained, however strong the reasons for it may be. This is what enabled him to maintain his most characteristic claim, even after the qualifiers he inserted in the Postscript: "The conversion experience that I have likened

to a gestalt switch remains, therefore, at the heart of the revolutionary process. Good reasons for choice provide motives for conversion and a climate in which it is more likely to occur. Translation may, in addition, provide points of entry for the neural reprogramming that, however inscrutable at this time, must underlie conversion. But neither good reasons nor translation constitute conversion, and it is that process we must explicate in order to understand an essential sort of scientific change."⁷

How is the transition to be explicated? Kuhn has only some hints to offer: "With respect to divergences of this sort, no set of choice criteria yet proposed is of any use. One can explain, as the historian characteristically does, why particular men made particular choices at particular times. But for that purpose one must go beyond the list of shared criteria to characteristics of the individuals who make the choice. One must, that is, deal with characteristics which vary from one scientist to another without thereby in the least jeopardizing their adherence to the canons that make science scientific."⁸

And he mentions such characteristics as previous experience as a scientist, philosophical views, personality differences. In the years since SSR appeared, sociologists of science have made much of these factors, often in ways that Kuhn himself would disavow. It was his stress on the role of these factors, he later remarked, that led critics to dub his views "subjectivist." They forgot his stress on the "shared criteria" that guide (but do not dictate) theory choice.⁹ I will take him at his word here, assuming that the rationality of theory choice in his account rests on the persistence of these criteria that enable theories to be compared and evaluated, relatively to one another, even when they are incommensurable.

2 | How Deep Do Revolutions Go?

Here we immediately encounter a difficulty. Do these criteria persist? Can they bridge paradigm differences? How deep, in short, do revolutions go? There is an ambiguity in Kuhn's response to this question. In a celebrated paragraph in SSR, he describes paradigm change as follows: "Like the choice between competing political institutions, that between competing paradigms proves to be a choice between incompatible modes of community life. Because it has that character, the choice is not and cannot be determined merely by the valuative procedures characteristic of normal science, for these depend in part upon a particular paradigm, and that paradigm is at issue. When paradigms enter, as they must, into a debate about paradigm choice, their role is necessarily circular. Each group uses its own paradigm to argue in that paradigm's defense."¹⁰

Since the evaluative procedures depend on the paradigm, and the paradigm itself is in question, there can be no agreed-upon way to adju-

dicating the choice between rival paradigms. Though he goes on to say that the resulting circularity does not *necessarily* undercut the arguments used, he concludes that the status of such arguments can at best be only that of persuasion. They "cannot be made logically or even probabilistically compelling for those who refuse to step into the circle. The premises and values shared by the two parties to a debate over paradigms are not sufficiently extensive for that."¹¹

What prevents the rival parties from agreeing as to which paradigm is the better, then, is in part the fact that the norms in terms of which this debate could be carried on are themselves part of the paradigm, so that there is no neutral methodological ground, or at least not enough to enable agreement to be reached. How important is this sort of "circularity" to Kuhn's account of the inability of either side in a paradigm debate to muster an entirely cogent argument in its own behalf? If a circularity in regard to evaluative procedures were to hold in general in such cases, then scientific revolutions *would* indeed seem to be the irrational, or at least minimally rational, affairs that Kuhn's critics take him to be saying they are. One way to find out is to direct attention to the examples he gives of scientific revolutions and ask what paradigm change amounts to in each of these cases.

When the question is put in this way, it is clear that there is a striking difference in the depth of the different changes classified by Kuhn as "revolutions." At one end of the spectrum is the Copernican revolution, the charting of which led him to the writing of *SSR* in the first place. At the other end would be, for example, the discovery of X rays. Somewhere in the middle might come the discovery of the oxygen theory of combustion.¹² We have a choice in some cases, it would seem, between saying that only a small part of the paradigm changed and saying that an entire paradigm changed but that the "paradigm" in this case comprised only a fraction of the beliefs, procedures, and so forth, of the scientists involved.

Take the case of X rays. Kuhn insists that their discovery did accomplish a revolution in his sense. Yet he recognizes that at first sight this episode scarcely seems to qualify. After all, no fundamental change of theory occurred. No troublesome anomalies were noted in advance. There was no prior crisis to signal that a revolution might be at hand. Why then, he asks, can we not regard the discovery of X rays as a simple extension of the range of electromagnetic phenomena? Because, he responds, it "violated deeply entrenched expectations . . . implicit in the design and interpretation of established laboratory procedures."¹³ The use of a particular apparatus "carries with it the assumption that only certain sorts of circumstances will arise." Roentgen's discovery "denied previously paradigmatic types of instrumentation their right to that title." That was sufficient, in his view, to constitute it a "revolution" in the sense in which he is proposing to use that term.

I will call this a shallow revolution because so much was left un-

touched by it. Electromagnetic theory was not replaced or even altered in any significant way. There were no challenges to accepted ways of assessing theory or to what counts as proper explanation. The textbooks, the sets of approved problem solutions, did not change much. What changed were the experimental procedures used in working with cathode-ray equipment and the expected outcomes of such work. And, of course, there were some important long-range implications for theory (as we now know). Such "revolutions" ought, it would seem, to be fairly frequent. Much would depend on how literally one should take the criteria Kuhn specifies as being the symptoms of impending revolution: previous awareness of anomaly and a resistance to a threatened change in procedures or categories.¹⁴

We are much more likely to think in terms of "revolution" in cases where one large-scale theory replaces another. Kuhn's favorite example is the replacement of phlogiston theory by the oxygen theory of combustion.¹⁵ It meant a reformulation of the entire field of chemistry, a new conceptual framework, a new set of problems. Another example he gives of this sort of intermediate revolution, as we might call it, is the discovery of the Leyden jar and the resulting emergence of "the first full paradigm for electricity."¹⁶ Prior to this discovery, Kuhn remarks, no single paradigm governed electrical research. A number of partial theories were applied, none of them entirely successful. The new conceptual framework enabled normal science to get under way, even though one-fluid and two-fluid theories were still in competition.

These changes involved the formulation of a new and more comprehensive theory. But they left more or less unchanged the epistemic principles governing the paradigm debate itself. Both sides would have agreed as to what counts as evidence, as to how claims should be tested. Or more accurately, to the extent that the scientists involved would have disagreed on these issues, their disagreements would not have been paradigm-dependent to any significant extent. So far as we can tell, Priestley and Lavoisier applied the same sorts of criteria to the assessment of theory, though they might not have attached the same weight to each criterion.

In Kuhn's favorite example of a scientific revolution, the Copernican one, this was, of course, not the case. This was a revolution of a much more fundamental sort because it involved a change in what counted as a good theory, in the procedures of justification themselves. It was not abrupt; indeed, it took a century and a half, from Copernicus's *De revolutionibus* to Newton's *Principia*, to consummate. And what made it revolutionary was not just the separation of Newtonian cosmology or Newtonian mechanics from their Aristotelian counterparts but the gradual transformation in the very idea of what constitutes valid evidence for a claim about the natural world, as well as in people's beliefs about how that world is ordered at the most fundamental level.¹⁷

It can thus be called a *deep* revolution, by contrast with the others described above. The Aristotelians and the Galileans totally disagreed as

to how agreement itself should be brought about. So did the Cartesians and the Newtonians. The Galileans made use of idealization, of measurement, of mathematics, in ways the Aristotelians believed were illegitimate. The Newtonians allowed a form of explanation that the Cartesians were quite sure was improper. The shift in paradigm here meant a radical shift in the methodology of paradigm debate itself. Paradigm replacement means something much more thoroughgoing in such a case.

Have there been other deep revolutions in the more recent history of natural science? Newton's success means the success of a methodology which is still roughly the methodology of natural science today. Perhaps only one deep revolution was needed to get us to what Kuhn calls "mature" science. The two major revolutions in the physics of our own century did not run quite so deep. But they *did* involve principles of natural order, that is, shared assumptions as to what count as acceptable ways of articulating physical process at its most basic level. In the quantum revolution, what separated Bohr and Einstein was not just a difference in theoretical perspective but a disagreement as to what counted as good science and why. Quantum theory, in its Copenhagen interpretation, came much closer to a deep paradigm replacement than it would have done in Einstein's way of taking it.

In the Postscript to SSR, Kuhn addressed the ambiguity of the notion of paradigm and proposed a new label. A disciplinary matrix is the answer to the question, "What does [a community of specialists] share that accounts for the relative fullness of their professional communication and the relative unanimity of their professional judgments?"¹⁸ Some of its principal components, he says, are symbolic generalizations, models of the underlying ontology of the field under investigation, concrete problem solutions, and the values governing theory appraisal.

It is clear, then, that for there to be a revolution in Kuhn's sense of the term this last component does not have to be at issue. Only in a deep revolution does one side challenge the other in regard to the appropriate methodology of theory assessment. When X rays were discovered, there was no dispute as to how their reality should be tested. When a Kuhnian revolution takes place, it is evidently not necessary that the entire paradigm should change. Only a part of the disciplinary matrix need be affected for there to be a sufficient change in worldview to qualify as "revolutionary." What 'revolutionary' means in practice is a change that falls outside the normal range of puzzle-solving techniques and whose resolution cannot, therefore, be brought about by the ordinary resources of the paradigm.

The implicit contrast is between puzzle solving, with its definitive ways of deciding whether a puzzle really *is* solved, and paradigm debate, where no such means of ready resolution exists. Whether so sharp a contrast is warranted by the actual practice of science may well be questioned. Decision between rival theories is an everyday affair in any active part of science. There may be an accepted general framework within which prob-

lems are formulated, but new data constantly pose challenges to older subtheories within that framework. This was the main issue dividing Kuhn and his Popperian critics in the late 1960s. It is clear in retrospect that there was merit on both sides of that dispute but that each was focusing on a particular aspect of scientific change to the exclusion of others.

The appraisal of rival theories within a paradigm is not a simple matter of puzzle solving. The history of high-energy physics over the past thirty years, for example, has seen one theory dispute after another. The notorious divisions at the moment among paleontologists about the causes of the Cretaceous extinction or between planetary physicists about the origin of the moon are only two of the more obvious reminders of the fact that deep-seated disagreement about the merits of alternative theories is a routine feature of science at its most "normal." As we have seen, Kuhn traced the roots of paradigm disagreement to two different sources: an "incommensurability" of a complex sort between two ways of looking at the world and a set of criteria for theory choice that function as values to be maximized rather than as an effective logic of decision. But this latter source of difference characterizes theory disputes generally and not just the more intractable ones that Kuhn terms paradigm disagreements. What we have here, I suspect, is a spectrum of different levels of intractability, not just a sharp dichotomy between revolutions and puzzle solutions. Nevertheless, Kuhn's dichotomy, though rather idealized, did serve to bring out in a forceful and dramatic way how complex, and how far from a simple matter of demonstration, the choice between theoretical alternatives ordinarily is.

3 | The Virtues of a Good Theory

What makes this choice a *rational* one for Kuhn, as we have seen, is the fact that scientists are guided by what they would regard as the virtues of a good theory. And there has been a certain constancy in that regard, according to him, across all but perhaps the deepest of revolutions: "I have implicitly assumed that, whatever their initial source, the criteria or values deployed in theory choice are fixed once and for all, unaffected by their transitions from one theory to another. Roughly speaking, but only roughly speaking, I take that to be the case. If the list of relevant values be kept short (I have mentioned five, not all independent) and if their specification is left vague, then such values as accuracy, scope, and fruitfulness are permanent attributes of science."¹⁹

This is a strong assertion indeed. Ironically, it is stronger than that now made by some of those who, like Laudan and Shapere, have chided Kuhn in the past for his subjectivism.²⁰ They argue that the values involved in theory choice are in no sense fixed; Shapere objects to any such claim as an objectionable form of essentialism. According to Laudan and Shap-

ere, these values themselves change gradually as theories change or are replaced. They change for *reasons*, they insist, these reasons functioning as some sort of higher-level arbitration. But there is no limit in principle as to how *much* they might change over time. To put this in a more direct way, there is no constraint on how different the criteria of a good theory might be in the science of the far future from those we rely on today, unlikely though a radical shift might be.²¹ In the original text of SSR, Kuhn proposed what sounds like a rather different view:

When paradigms change, there are usually significant shifts in the criteria determining the legitimacy both of problems and of proposed solutions. . . . [This is] why the choice between competing paradigms regularly raises questions that cannot be resolved by the criteria of normal science. To the extent, as significant as it is incomplete, that two scientific schools disagree about what is a problem and what a solution, they will inevitably talk through each other when debating the relative merits of their respective paradigms. In the partially circular arguments that regularly result, each paradigm will be shown to satisfy more or less the criteria that it dictates for itself and to fall short of a few of those dictated by its opponent.²²

The criteria governing theory choice are described here as strongly paradigm-dependent and thus as suffering "significant shifts" from one paradigm to the next. The resulting partial circularity in paradigm assessment leads rival scientists to "talk through each other." This was the theme, of course, that Paul Feyerabend picked up on. One can see how severely it limits the notion that there are "good reasons" for paradigm change. Here, then, is a clear instance of how Kuhn's later construals soften the radical overtones of the earlier work.

Kuhn does not hesitate to speak of the values involved in theory appraisal as "permanent attributes of science." He allows that the manner in which these values are understood and the relative weights attached to them have changed in the past and may change again in the future. But he wants to emphasize that these changes at the metalevel tend to be slower and smaller in scale than the changes that can occur at the level of theory:

If such value changes had occurred as rapidly or been as complete as the theory changes to which they related, then theory choice would be value choice, and neither could provide justification for the other. But, historically, value change is ordinarily a belated and largely unconscious concomitant of theory choice, and the former's magnitude is regularly smaller than the latter's. For the functions I have here ascribed to values, such relative stability provides a sufficient basis. The existence of a feedback loop through which theory change affects the values which led to that change does not make the decision process circular in any damaging sense.²³

One would need, however, to know just how and why changes in theory bring about changes at the metalevel of theory assessment in order to judge how large these latter changes might become without undermining the claim that a rational choice is being made. Is the "relative stability" of the criteria governing theory choice a contingent historical finding, or is it a necessary feature of any activity claiming the title of science? There are suggestions of both views in the passage I have just quoted. Historically, these values have in fact been stable, Kuhn remarks. But he adds that if they were not, if one had to choose the criteria of choice themselves in the act of choosing between theories, there would be no fulcrum. The process would lack justification; it would be circular in a way that would be damaging to its claim to qualify as science.

The presumption appears to be that *really* deep revolutions do not occur, that is, revolutions where there is no sharing of epistemic values between one paradigm and the other. Kuhn allows that large-scale theory change may involve smaller-scale changes in the values believed to be appropriate to theory appraisal. In such cases, adoption of the new paradigm carries with it adoption of a somewhat different "rationality" at the metalevel. The advantages of the new theory are so marked, in terms of a minimal level of shared values, that a shift in the values themselves is ultimately taken to be warranted. This, it can be argued, is what happened in the seventeenth century as the balance shifted between Aristotelians and Galileans. Galileo set out to undermine Aristotle's physics in its *own* terms first and then to present an alternative that, in terms of consistency, empirical adequacy, and future potential, could claim a definite advantage, even in terms of criteria the Aristotelian might be brought to admit. That, at any rate, would be the grounds, in Kuhn's perspective, for regarding the Scientific Revolution as a "rational" shift in the way in which natural science was carried on.

In a recent essay Kuhn argues that we learn to use the term 'science' in conjunction with a cluster of other terms like 'art', 'medicine', 'philosophy'. To know what science is, is to know how it relates to these other activities.²⁴ Identifying an activity as scientific is to single out "such dimensions as accuracy, beauty, predictive power, normativeness, generality, and so on. Though a given sample of activity can be referred to under many descriptions, only those cast in this vocabulary of disciplinary characteristics permit its identification as, say, science; for that vocabulary alone can locate the activity close to other scientific disciplines and at a distance from disciplines other than science. That position, in turn, is a necessary property of all referents of the modern term, 'science.'"²⁵

He immediately qualifies this last very strong claim by noting that not every activity that qualifies as "scientific" need be predictive, not all need be experimental, and so forth. And there is no sharp line of demarcation between science and nonscience. Nonetheless, there is a well-defined cluster of values whose pursuit marks off scientific from other activities in a

relatively unambiguous way and that gives the term 'science' the position it occupies in the "semantic field." This marking off is not a mere matter of convention. The taxonomy of disciplines has developed in an empirical way; a real learning has taken place. If someone were to deny the rationality of learning from experience, we would not know what he or she is trying to say. One cannot, he maintains, further *justify* the norms for rational theory choice. He cites C. G. Hempel to the effect that this inability is a testimony to our continuing failure to solve the classical problem of induction.²⁶

Kuhn rests his case, then, both for the rationality of science and for its distinctiveness as a human activity mainly on the values governing theory choice in science. But he does not chronicle their history, disentangle them from one another except in a cursory way, or inquire in any detail into how and why they have changed in the ways they have. Many of these variations, he remarks, "have been associated with particular changes in scientific theory. Though the experience of scientists provides no philosophical justification for the values they deploy (such justification would solve the problem of induction), those values are in part learned from that experience, and they evolve with it."²⁷

But what justification other than the experience of scientists is *needed* to justify the values they deploy? Kuhn has, I suspect, altogether too lofty a view of what "philosophical" justification might amount to. And he has too readily allowed himself to be intimidated by that most dire of philosophers' threats: "That *can't* be right: if it were, it would solve the problem of induction." My own guess is that attention to the role of values in theory appraisal might well dissolve the problem Hume bequeathed us about the grounds for inductive inference. But whether that be true or not, the criteria employed by scientists in theory evaluation enjoy whatever sanction is appropriate to something learned in, and tested by, experience.

4 | How Might Epistemic Values Be Validated?

Suppose a scientist were to doubt whether a particular value, say simplicity, is really a desideratum in a practical situation of theory choice facing him or her. The rationality of the choice depends, presumably, on what sort of answer can be given to this kind of question. Two different sorts of answers suggest themselves. One is to look at the track record and decide how good a guide simplicity has proved to be in the past. (There are obvious problems about how the criterion itself is to be understood, but I will bracket these for the moment.) A quite different sort of response would be that simplicity is clearly a desideratum of theory because —, where we fill the blank with a reason why on the face of it, a simple theory is more likely to be a good theory (if indeed one *can* find a convincing

reason). Both of these responses would, of course, need further clarification before they could begin to carry any conviction.

First, what does it mean to ask how good a guide simplicity has been in the past? Guide to what? Some kind of ordering of means and ends is clearly needed here. Some of the values we have been talking about seem to function as goals of the scientific enterprise itself: predictive accuracy (empirical adequacy) and explanatory power are the most obvious candidates. One can trace each of these goals back a very long way in human history. In some sense, they may be as old as humanity itself. The story of how they developed in the ancient world, how the skills of prediction came to be prized in many domains, how explanatory accounts of natural process came to be constructed, is a familiar one. Less familiar is the realization that these goals were not linked together in any organic way at the beginning. Indeed, they were long considered antithetical in the domain of astronomy, the most highly developed part of the knowledge of nature in early times. One of the consequences, perhaps the most important consequence, of the Copernican revolution was to show that they *are* compatible, that they can be successfully blended. This was an empirical discovery about the sort of universe we live in. It was something we *learned* and that now we *know*.

Each of these goals has come to be considered valuable in its own right, an end in itself.²⁸ An activity that gives us accurate knowledge of the world we live in and consequently power over its processes can come to seem worthwhile for all sorts of reasons. An activity that allows us to understand natural process, that allows our imaginations to reach out to realms inaccessible to our senses, holds immediate attraction. What it is to understand will, of course, shift as the principles of natural order themselves shift. So this goal of explaining lacks the definiteness of the goal of predicting; as theory changes, so will the contours of what counts as explaining.

Much more would have to be said about all this, but I am going to press on to make my main point.²⁹ Other epistemic values serve as *means* to these ends; they help to identify theories more likely to predict well or to explain. Some of these are quite general and would apply to any epistemic activity. Logical consistency (absence of contradiction) and compatibility with other accepted knowledge claims would be among these. They are obviously not goals in themselves; they would not motivate us to carry on an activity in the first place. But we have found that these values are worth taking seriously as *means*. Or should I say, it has always been obvious that we must not neglect them, if it is knowledge we are seeking?

Other values are more specific to science, for example, fertility, unifying power, and coherence (i.e., absence of ad hoc features). Once again, these are clearly not primary goals. They are not so much deliberately aimed at as esteemed, when present. And they are esteemed not in them-

selves but because they have proved to be the marks of a "good" theory, a theory that will serve well in prediction and explanation. A long story could be told about this, beginning with Kepler, Boyle, and Huygens and working through Herschel, Whewell, and a legion of others who have drawn attention to the significance of these three virtues.

Once again, the story is an ambiguous one: it can be told in two quite different ways. According to one way of telling it, these values can be shown to have played a positive historical role in theory choice; we have gradually learned to trust them as clues. According to the other, a series of acute thinkers (some of the most prominent of them listed above) have realized that these values *ought* to serve as indicators of a good theory. These are what one would *expect* a priori from a theory that purported to predict accurately and explain correctly. When Kepler and Boyle drew attention to the importance of such criteria, it was not to point to their efficacy in the earlier history of natural philosophy but to recommend them on general epistemic grounds.³⁰

The question of how to validate the values that customarily guide scientific theory choice can now be addressed more directly. The goals of predictive accuracy (empirical adequacy) and explanatory power serve to define the activity of science itself, in part at least. If, as Kuhn notes, one relinquishes the goal of producing an accurate account of natural regularity, the activity one is engaged in may be worthwhile, but it is not science.³¹ The notion of epistemic justification does not directly apply to the goals themselves. One might ask, of course, whether the pursuit of these goals is justifiable on *moral* grounds. Or one might ask, as a means of determining whether effort expended on them is worthwhile, whether the goals are in fact attainable. We have learned that in general they *are* attainable. This is something one could not have known a priori. And we have learned much about the *methods* that have to be followed for theory construction to get under way, methods of experiment, of conceptual idealization, of mathematical formulation, and the rest. All of this had to be *learned*, and no doubt there is still much to discover in this regard.

The other values, being instrumental, are justified when it is shown that they serve as means to the ends defined by the primary goals. And this, as we have seen, can be done in two ways: by an appeal to what we have learned from the actual practice of science or by an analysis in epistemological terms of the aims of theory and what, in consequence, the marks of a good theory should be. Ideally, both ways need to be followed, each serving as check for the other. The appeal to historical practice works not so much as a testimony to what values have actually guided scientists in their theory choices but as a finding that reliance on certain values has *in fact* served the primary goals of science. Might it cease to?

This is the Humean echo that seems to worry Kuhn so much. One might respond, as he does, that learning from experience is part of what it is to be rational. We cannot *demonstrate* that experience will continue

to serve as a reliable guide. But demonstration is not what is called for. Kuhn has done more than anyone else, perhaps, to show that rational theory choice does not require the cogency of demonstration. We know that the predictive powers of natural science have enormously increased, and we know something of the theory characteristics that have served to promote this expansion. No future development could, so far as I can see, lead us to deny these knowledge claims, which rest not just on a perception of past regularities but on an understanding, partial at least, of why these regularities took the course they did. We can, and almost surely will, learn more about what to look for in a good theory. But no further evidence seems to be needed to show that coherence in a theory is a value to be sought, so that, other things being equal, a more coherent theory is to be preferred to a less coherent one.

5 | Rationality without Realism?

Over the years since SSR appeared, Kuhn has, as we have seen, become more and more explicit about the basic rationality that underlies theory choice in science. It is a complex rationality with many components, allowing much latitude for difference among the defenders of different theories. But it has remained relatively invariant since the deep revolution that brought it into clear focus in the seventeenth century. One might almost speak of a *convergence* here. Kuhn clearly believes that scientists have a pretty good grip on the values that *ought* to guide the appraisal of rival theories, and that this grip has improved as it has been tested against a wider and wider variety of circumstances.

But he has not softened his stance in regard to the truth character of theories in the least. In a well-known passage in the *Postscript*, he insists that the only sort of progress that science exhibits is in puzzle solving: later theories solve more puzzles than earlier ones, or (to put this in a different idiom) they predict better. But there is, he insists, "no coherent direction of ontological development"; there is no reason to think that successive theories approximate more and more closely to the truth.³² "The notion of a match between the ontology of a theory and its 'real' counterpart in nature now seems to me illusive in principle."³³ Kuhn thus rejects in a most emphatic way the traditional realist view that the explanatory success of a theory gives reason to believe that entities like those postulated by the theory exist, i.e., that the theory is at least approximately true.

He does not argue for this position in SSR, aside from a remark about Einstein's physics being closer in some respects to Aristotle's than to Newton's. But it is clear what the grounds for it are in his mind: the incommensurability of successive paradigms implies a discontinuity between

their ontologies. By separating the issues of comparability and commensurability, he believes he can retain a more or less traditional view in regard to the former while adopting an instrumentalist one in regard to the latter. The radical challenge of SSR is directed not at rationality but at realism. The implications of the familiar Kuhnian themes of holism and paradigm replacement are now seen to be more significant for the debate about realism than for the issue of scientific rationality, on which they had so great an initial impact.

Kuhn's influence on the burgeoning antirealism of the last two decades can scarcely be overestimated. His views on theory change, on problems about the continuity of reference, are reflected in the work of such notable critics of realism as Arthur Fine, Bas van Fraassen, and especially Larry Laudan.³⁴ Kuhn's own emphasis on science as a puzzle-solving enterprise would lead one to interpret him in an instrumentalist manner. At this point I am obviously not going to open a full-scale debate on realism versus instrumentalism.³⁵ But I would like to pull out one thread from that notorious tangle. Kuhn's way of securing scientific rationality by focusing on the values proper to theory choice might well have led him (I argue) to a more sympathetic appreciation of realism. I am not saying that rationality and realism are all of a piece, that to defend one is to commit oneself to the other. Most of the current critics of realism would be emphatic in their defense of the overall rationality of scientific change. But a closer study of the values to which Kuhn so effectively drew attention should, to my mind, raise a serious question about the adequacy of an instrumentalist construal of the puzzle-solving metaphor. If such a construal is adopted, it is hard to make sense of those many episodes in the history of science where values other than mere predictive accuracy played a decisive role in the choice between theories.

To show this, I will focus on a case history from Kuhn's own earlier work, *The Copernican Revolution*. At issue are the relative merits of the Ptolemaic and the Copernican systems prior to Galileo's work. Kuhn points out that there was little to choose between the two on the score of predictive accuracy. "Judged on purely practical grounds," he concludes, "the Copernican system was a failure; it was neither more accurate nor significantly simpler than its Ptolemaic predecessors."³⁶ Yet it persuaded some of the best astronomers of the time. And it was they who ultimately produced the "simple and accurate" account that carried the day. How *did* it persuade them? In Kuhn's view, "The real appeal of sun-centered astronomy was aesthetic rather than pragmatic. To astronomers the initial choice between Copernicus' system and Ptolemy's could only be a matter of taste, and matters of taste are the most difficult of all to define or debate."³⁷

But such matters cannot be regarded as unimportant, he goes on, as the success of the Copernican Revolution itself testifies. Whatever it was that persuaded so many of those most skilled in astronomy to make what

we would now regard as the right step obviously must be looked at with care. Those who were equipped "to discern geometric harmonies" obviously found "a new neatness and harmony" in the heliocentric system. What Copernicus offered was "a new and aesthetic harmony" that somehow carried conviction in the right quarters.

But now let us see how Copernicus's own argument went, in the crucial chapter 10 of book 1 of *De revolutionibus*. He points to two different sorts of clues. First, the heliocentric model allows one to specify the order of the planets outward from the central body in an unequivocal way, which Ptolemy's model could not do. Furthermore, the Copernican model has the planetary periods increase as one moves outward from the sun, just as one would expect. What Copernicus claims to discover in the new way of ordering the planets is a "clear bond of harmony," "an admirable symmetry." But why should this carry conviction, especially since (as Kuhn emphasizes) Copernicus in the end had to retain an inelegant and far from harmonious-seeming tangle of epicycles?

He had stronger arguments. The heliocentric model could *explain*, that is, provide the *cause* of, a whole series of features of the planetary motions that Ptolemy simply had to postulate as given, as inexplicable in their own right. For example, even in ancient times it had been suggested that Venus and Mercury appear to have the sun as their center of rotation, since, unlike the other planets, they accompany the sun in its motion across our sky. Or again, it had long been noted that the superior planets (Mars, Jupiter, Saturn) are at their brightest when in opposition (rising together in the evening or setting together in the morning). Assuming that brightness is a measure of relative distance, this is explained if we are viewing the planetary motions from a body that itself is orbiting the sun as center. This "proves," Copernicus somewhat optimistically concludes, that the center of motion of the superior planets is the same as that of the inferior planets, namely the sun.

Kuhn comments that it does "not actually prove a thing. The Ptolemaic system explains these phenomena as completely as the Copernican," although the latter can be said to be "more natural."³⁸ Here I must disagree. The Ptolemaic system does not *explain* the phenomena mentioned above at all. Ptolemy is forced to postulate that the center of the epicycle for both Venus and Mercury always lies on the line joining the earth and sun. Kuhn says that in this way Ptolemy "accounts for" this feature of their motions. But this is surely not *accounting for* in the sense of explaining. Kuhn evidently equates prediction and explanation in these passages, not an unusual assumption at the time his book was written.

But he allows that Copernicus gives a "far more natural" account than does Ptolemy. Why? And what does 'natural' mean in the lexicon of an instrumentalist? Ptolemy's restriction on the deferent radii swept out by Venus and Mercury "is an 'extra' device, an *ad hoc* addition,"³⁹ one that Copernicus can discard. Kuhn is surely on the right track here. But this

is *not* an aesthetic argument, an appeal to taste. Copernicus himself makes the genre to which it belongs quite clear. He says that [he] is able to assign the cause of these features of the planetary motions, whereas Ptolemy is not. There is no reason in Ptolemy's system for them, other than the mere need to get the predictions right. They are, as Kuhn himself says, *ad hoc*.

Copernicus gives another set of arguments based on the retrograde motions. Their relative size and frequency from one planet to another and the lack of any such motions on the part of the sun and moon are exactly what one would be led to expect in a system where we are observing the motions from the third planet and the moon is not a true planet but a satellite of earth. Later, in the *Mysterium cosmographicum*, Kepler developed these arguments more fully and added some of his own, for example, the striking fact that in the Ptolemaic model, the period of rotation for each planet on either the deferent or the epicycle circle is exactly one year, something which seemed like an extraordinary piece of adjustment, especially since Ptolemy took the planets to be dynamically independent of one another. Kepler is clear that the issue here is one of causal explanation; one of the systems can provide such an explanation, the other cannot. He is also clear that the criterion of prediction alone will not be enough to decide in all cases between two rival accounts of the planetary motions and thus that a different genre of argument (he calls it "physical") is needed.⁴⁰ This he urged as a refutation of the instrumentalism of his opponent, Ursus.

The competition may have been neutral between Ptolemy and Copernicus where *prediction* of planetary motions was concerned, but the two systems were quite unequal as *explanation*. No better illustration could be found of the distinction between these two concepts, and of the consequent importance of criteria of theory appraisal other than that of predictive or descriptive accuracy. Copernicus's criterion of "naturalness," the elimination of *ad hoc* features, the virtue that might today be called coherence, is not aesthetic; it is epistemic. He is not just appealing to his reader's taste, or sense of elegance. He is not assuming that the simpler, the more beautiful, models are more likely to be true. He is saying that a theory that makes causal sense of a whole series of features of the planetary motions is more likely to be true than one that leaves these features unexplained.

Copernicus and those who followed him believed that they had good arguments for the reality of the earth's motion around the sun. They sometimes overstated the force of those arguments, to be sure, using terms like 'proof' and 'demonstration'. The natural philosophers of the day were not yet accustomed to the weaker notions of likelihood and probability. Galileo found, to his cost, that he had to speak in terms of demonstration if his claims for the Copernican system were to be taken seriously. He did not have a demonstration, but from our perspective, he called effectively

on the criterion of coherence in his critique of the geostatic alternative, just as Copernicus had ~~earlier~~ done.

As we look back on those debates, we are ready to allow that the coherence arguments of Copernicus and Galileo *did* carry force, that they *did* give a motive for accepting the new heliocentric model as true. And their force came from something other than predictive advantage. Kuhn's point in regard to theory assessment, one that became clearer in his successive formulations of it, was that the different theory values were not reducible to one another, and hence that no simple algorithm, no logic of confirmation such as the logical positivists had sought, underlay real-life theory decision. What I have tried to do here is to carry this insight further and to note the special epistemic weight carried by certain of these values. Besides coherence, one could make similar cases for fertility and unifying power. It is hard to make sense of the role played by these values if one adopts the instrumentalist standpoint that Kuhn feels compelled to advocate.

The case for scientific realism rests in large part on these "superempirical" values. That is, when we ask about a particular theory, how likely is it that it is true (correlatively, how likely is it that something like the explanatory entities it postulates actually exist), it is to these virtues that we are inclined to turn. To say that a theory simply "saves the phenomena," though this carries *some* epistemic weight, leaves open the suspicion of its being ad hoc. If a theory be thought of simply as an hypothetico-deductive device, it would seem plausible to suppose that other devices might account as well or better for the phenomena to be explained. It is only when the *temporal* dimension is added, when a theory is evaluated in a historical context, when its success in unifying domains over time or in predicting new sorts of phenomena are taken into account, that conviction begins to emerge. Theories are not assessed simply as predictors; they are not confirmed purely by the enumeration of consequences.

My conclusion is that the diversity of the expectations scientists hold up for their theories argues not only for the tentative character of theory choice, Kuhn's original point, but also for its properly epistemic character. This leaves us, of course, with a problem: *how* can the difficulties in regard to incommensurability be reconciled with the epistemic force of such arguments as that of Copernicus? Kuhn emphasized the discontinuities of language across theory change so strongly that he left no room for the possibility of convergence, for the possibility that the theories of the paleontologists of today, for example, not only solve more puzzles than those of yesteryear but also tell us, with high degree of likelihood, what actually happened at distant epochs in the earth's past.

The Kuhnian heritage is thus a curiously divided one. Kuhn wanted to maintain the rational character of theory choice in science while denying the epistemic character of the theory chosen. The consequent

tensions are, of course, familiar to every reader of current philosophy of science. Thirty years later, *The Structure of Scientific Revolutions* still leaves us with an agenda.

■ | Notes

1. [Hereafter "Objectivity," from] *The Essential Tension* (ET), pp. 320–339 [reprinted this volume, pp. 102–18]. In his effort to ward off the charge of subjectivism, Kuhn might also have pointed to "The Function of Measurement in Modern Physical Science" (*Isis* 52 [1961]: 161–190; reprinted in ET, pp. 178–224), which appeared before SSR and whose theme was that "measurement can be an immensely powerful weapon in the battle between two theories" (ET, p. 211), that "the comparison of numerical predictions . . . has proved particularly successful in bringing scientific controversies to a close" (ET, p. 213). Or he could have recalled an even earlier paper, "The Essential Tension" (*The Third University of Utah Research Conference on the Identification of Scientific Talent*, ed. C. W. Taylor [Salt Lake City: University of Utah Press, 1959], 162–174; reprinted in ET, pp. 225–239), whose title referred to the opposition between the themes of tradition and innovation in science and which argued that it is the very effort to work within a tightly construed tradition that leads eventually to the recognition of anomalies that in turn prepares the way for revolution (ET, p. 234). One further paper that Kuhn might have called on was "A Function for Thought Experiments" (*L'Aventure de la science, Mélanges Alexandre Koyré* [Paris: Hermann, 1964], vol. 2, pp. 307–334; reprinted in ET, pp. 240–265), which describes how failures of expectation induce the crisis that is the usual prelude to paradigm change (ET, p. 263).
2. "Objectivity," p. 320 [p. 102, above].
3. *The Structure of Scientific Revolutions*, 2nd ed. (Chicago: University of Chicago Press, 1970), p. 199.
4. "Metaphor in Science," in *Metaphor and Thought*, ed. Andrew Ortony (Cambridge: Cambridge University Press, 1979), 409–419; see p. 416.
5. In a recent essay Kuhn distinguishes between translation and interpretation and shows how communication can occur even where languages are incommensurable ("Commensurability, Comparability, Communicability," *PSA 1982* [Philosophy of Science Association], 1983: 669–688). In a comment Philip Kitcher remarks that Kuhn, in his later readings of SSR, has progressively weakened the dramatic doctrine of the original work in ways, be it said, of which Kitcher approves ("Implications of Incommensurability," *PSA 1982*, 1983: 689–703).
6. SSR, p. 204.
7. SSR, p. 204.
8. "Objectivity," p. 324 [p. 106, above].
9. "Objectivity," p. 325 [p. 106, above].
10. SSR, p. 94 [p. 88, above].

11. SSR, p. 94 [p. 88, above].
12. In SSR Kuhn himself distinguishes between "major paradigm changes, like those attributable to Copernicus and Lavoisier," and "the far smaller ones associated with the assimilation of a new sort of phenomenon, like oxygen or X-rays" (p. 92) [p. 86, above].
13. SSR, p. 58.
14. SSR, p. 62. It is not clear to me that the discovery of X rays satisfies either of these criteria in any other than a minimal way.
15. SSR, p. 199.
16. SSR, p. 62. There might be some question as to whether, in fact, a single theory of electricity did emerge at this time. But that is not to the point of my inquiry.
17. I have worked out this theme in some detail in my "Conceptions of Science in the Scientific Revolution," in *Reappraisals of the Scientific Revolution*, ed. David Lindberg and Robert Westman (Cambridge: Cambridge University Press, 1990).
18. SSR, p. 182.
19. "Objectivity," p. 335 [p. 114, above].
20. See Larry Laudan, *Science and Values* (Berkeley: University of California Press, 1984) [Ch. 4 of *Science and Values* is reprinted on pp. 139–69, this volume]; Dudley Shapere, *Reason and the Search for Knowledge* (Dordrecht: Reidel, 1984). I have discussed the ironies of this particular divergence more fully in "The Shaping of Scientific Rationality," in *Construction and Constraint*, ed. E. McMullin (Notre Dame: University of Notre Dame Press, 1988), pp. 1–47.
21. Nicholas Rescher defends a somewhat similar position in regard to how different from ours the "science" carried on by the inhabitants of a distant planet might be: "Science as we have it—the only 'science' that we ourselves know—is a specifically human artifact that must be expected to reflect in significant degree the particular characteristics of its makers. Consequently, the prospect that an alien 'science'-possessing civilization has a *science* that we could acknowledge (if sufficiently informed) as representing the same general line of inquiry as that in which we ourselves are engaged seems extremely implausible" ("Extraterrestrial Science," *Philosophia Naturalis* 21 [1984]: 400–424; see p. 413).
22. SSR, pp. 109–110 [p. 100, above].
23. "Objectivity," p. 336 [p. 115, above].
24. "Rationality and Theory Choice," *Journal of Philosophy* 80 (1983): 563–570; see p. 567.
25. *Ibid.*, p. 568.
26. C. G. Hempel, "Valuation and Objectivity in Science," in *Physics, Philosophy, and Psychoanalysis*, ed. R. S. Cohen and L. Laudan (Dordrecht: Reidel, 1983), 73–100.
27. "Objectivity," p. 335 [p. 114, above].

28. See my "Values in Science," *PSA* 1982, 1983: 3–25.
29. The story sketched so lightly here is told in much more detail in my "Goals of Natural Science," *Proceedings of the American Philosophical Association* 58 (1984): 37–64.
30. For a fuller historical treatment, see my "Conceptions of Science in the Scientific Revolution."
31. "Rationality and Theory Choice," p. 569.
32. *SSR*, p. 206.
33. *SSR*, p. 206.
34. Laudan's much-quoted essay, "A Confutation of Convergent Realism," in *Scientific Realism*, ed. J. Leplin (Berkeley: University of California Press, 1984), 218–249 [an edited version is reprinted on pp. 1114–35, this volume], presents in detail the sort of arguments that Kuhn would need to support his own rejection of convergence.
35. See my "Case for Scientific Realism," in *Scientific Realism*, ed. J. Leplin, 8–40, and "Selective Anti-realism," *Philosophical Studies* 61 (1991): 97–108.
36. *The Copernican Revolution* (New York: Random House, 1957), p. 171.
37. *Ibid.*, p. 172.
38. *Ibid.*, p. 178.
39. *Ibid.*, p. 172.
40. Kepler's clearest treatment of this issue will be found in the *Apologia Tychonis contra Ursum* (1600). See Nicholas Jardine's translation of this work in *The Birth of History and Philosophy of Science* (Cambridge: Cambridge University Press, 1984). Michael Gardner extracts a "Kepler principle" to the effect that it counts in favor of the realistic acceptance of a theory if it explains facts that competing theories merely postulate. See "Realism and Instrumentalism in Pre-Newtonian Astronomy," in *Testing Scientific Theories*, ed. John Earman (Minneapolis: University of Minnesota Press, 1983), 201–265; p. 256.

Dissecting the Holist Picture of Scientific Change

It is now more than twenty years since the appearance of Thomas Kuhn's *The Structure of Scientific Revolutions*. For many of us entering the field two decades ago, that book made a powerful difference. Not because we fully understood it; still less because we became converts to it. It mattered, rather, because it posed in a particularly vivid form some direct challenges to the empiricism we were learning from the likes of Hempel, Nagel, Popper, and Carnap.

Philosophers of science of that era had no doubts about whom and what the book was attacking. If Kuhn was right, all the then reigning methodological orthodoxies were simply wrong. It was a good deal less clear what Kuhn's positive message amounted to, and not entirely because many of Kuhn's philosophical readers were too shocked to read him carefully. Was he saying that theories were really and always incommensurable so that rival scientists invariably misunderstood one another, or did he mean it when he said that the problem-solving abilities of rival theories could be objectively compared? Did he really believe that accepting a new theory was a "conversion experience," subject only to the Gestalt-like exigencies of the religious life? In the first wave of reaction to Kuhn's bombshell, answers to such questions were not easy to find.

Since 1962 most of Kuhn's philosophical writings have been devoted to clearing up some of the ambiguities and confusions generated by the language of the first edition of *The Structure of Scientific Revolutions*. By and large, Kuhn's message has been an ameliorative and conciliatory one, to such an extent that some passages in his later writings make him sound like a closet positivist. More than one commentator has accused the later Kuhn of taking back much of what made his message interesting and provocative in the first place.¹

But that is not entirely fair, for if many of Kuhn's clarifications have indeed taken the sting out of what we once thought Kuhn's position was, there are several issues about which the later Kuhn is both clear *and* controversial. Significantly, several of those are central to the themes of this essay. Because they are, I want to use Kuhn's work as a stalking-horse to show how the features of the reticulational model . . . can be used to produce a more satisfactory account than Kuhn offers of scientific debate in particular and scientific change in general.

Kuhn, then, will be my immediate target, but I would be less than candid if I did not quickly add that the views I discuss here have spread considerably beyond the Kuhnian corpus. To some degree, almost all of us who wrote about scientific change in the 1970s (present company included) fell prey to some of the confusions I describe. In trying to characterize the mechanisms of theory change, we have tended to lapse into sloppy language for describing change. However, because Kuhn's is the best-known account of scientific change, and because Kuhn most overtly makes several of the mistakes I want to discuss, this chapter focuses chiefly on his views. Similar criticisms can be raised with varying degrees of severity against authors as diverse as Foucault, Lakatos, Toulmin, Holton, and Laudan.

■ | Kuhn on the Units of Scientific Change

It is notorious that the key Kuhnian concept of a paradigm is multiply ambiguous. Among its most central meanings are the following three: First and foremost, a paradigm offers a conceptual framework for classifying and explaining natural objects. That is, it specifies in a generic way the sorts of entities which are thought to populate a certain domain of experience and it sketches out how those entities generally interact. In short, every paradigm will make certain claims about what populates the world. Such ontological claims mark that paradigm off from others, since each paradigm is thought to postulate entities and modes of interaction which differentiate it from other paradigms. Second, a paradigm will specify the appropriate methods, techniques, and tools of inquiry for studying the objects in the relevant domain of application. Just as different paradigms have different ontologies, so they involve substantially different methodologies. (Consider, for instance, the very different methods of research and theory evaluation associated with behaviorism and cognitive psychology respectively.) These methodological commitments are persistent ones, and they characterize the paradigm throughout its history. Finally, the proponents of different paradigms will, according to Kuhn, espouse different sets of cognitive goals or ideals. Although the partisans of two paradigms may

(and usually do) share some aims in common, Kuhn insists that the goals are not fully overlapping between followers of rival paradigms. Indeed, to accept a paradigm is, for Kuhn, to subscribe to a complex of cognitive values which the proponents of no other paradigm accept fully.

Paradigm change, on this account, clearly represents a break of great magnitude. To trade in one paradigm for another is to involve oneself in changes at each of the three levels [of paradigms]. We give up one ontology for another, one methodology for another, and one set of cognitive goals for another. Moreover, according to Kuhn, this change is *simultaneous* rather than *sequential*. It is worth observing in passing that, for all Kuhn's vitriol about the impoverishment of older models of scientific rationality, there are several quite striking similarities between the classical version of the hierarchical model and Kuhn's alternative to it. Both lay central stress on the justificatory interactions between claims at the factual, methodological, and axiological levels. Both emphasize the centrality of values and standards as providing criteria of choice between rival views lower in the hierarchy. Where Kuhn breaks, and breaks radically, with the tradition is in his insistence that rationality must be relativized to choices within a paradigm rather than choices between paradigms. Whereas the older account of the hierarchical model had generally supposed that core axiological and methodological commitments would typically be common property across the sciences of an epoch, Kuhn asserts that there are methodological and axiological discrepancies between any two paradigms. Indeed (as we shall see below), one of the core failings of Kuhn's position is that it so fully internalizes the classical hierarchical approach that, whenever the latter breaks down (as it certainly does in grappling with interparadigmatic debate, or any other sort of disagreement involving conflicting goals), Kuhn's approach has nothing more to offer concerning the possibility of rational choices.²

For now, however, the immediate point to stress is that Kuhn portrays paradigm changes in ways that make them seem to be abrupt and global ruptures in the life of a scientific community. So great is this supposed transition that several of Kuhn's critics have charged that, despite Kuhn's proclaimed intentions to the contrary, his analysis inevitably turns scientific change into a nonrational or irrational process. In part, but only in part, it is Kuhn's infelicitous terminology that produces this impression. Notoriously, he speaks of the acceptance of a new paradigm as a "conversion experience,"³ conjuring up a picture of the scientific revolutionary as a born-again Christian, long on zeal and short on argument. At other times he likens paradigm change to an "irreversible Gestalt-shift."⁴ Less metaphorically, he claims that there is never a point at which it is "unreasonable" to hold onto an old paradigm rather than to accept a new one.⁵ Such language does not encourage one to imagine that paradigm change is exactly the result of a careful and deliberate weighing-up of the

respective strengths of rival contenders. But impressions based on some of Kuhn's more lurid language can probably be rectified by cleaning up some of the vocabulary of *The Structure of Scientific Revolutions*, a task on which Kuhn has been embarked more or less since the book first appeared.⁶ No changes of terminology, however, will alter the fact that some central features of Kuhn's model of science raise serious roadblocks to a rational analysis of scientific change. The bulk of this chapter is devoted to examining some of those impedimenta. Before we turn to that examination, however, I want to stress early on that my complaint with Kuhn is not merely that he has failed to give any normatively robust or rational account of theory change, serious as that failing is. As I show below, he has failed even at the descriptive or narrative task of offering an accurate story about the manner in which large-scale changes of scientific allegiance occur.

But there is a yet more fundamental respect in which Kuhn's approach presents obstacles to an understanding of the dynamics of theory change. Specifically, by insisting that individual paradigms have an integral and static character—that changes takes place only between, rather than within, paradigms—Kuhn has missed the single feature of science which promises to mediate and rationalize the transition from one world view or paradigm to another. Kuhn's various writings on this subject leave the reader in no doubt that he thinks the parts of a paradigm go together as an inseparable package. As he puts it in *The Structure of Scientific Revolutions*, "In learning a paradigm the scientist acquires theory, methods, and standards together, usually in an *inextricable mixture*."⁷ This theme, of the inextricable and inseparable ingredients of a paradigm, is a persistent one in Kuhn's work. One key aim of this chapter is to show how drastically we need to alter Kuhn's views about how tightly the pieces of a paradigm's puzzle fit together before we can expect to understand how paradigmlike change occurs.

LOOSENING UP THE FIT

Without too heavy an element of caricature, we can describe world-view models such as Kuhn's along the following lines: one group or faction in the scientific community accepts a particular "big picture." That requires acquiescence in a certain ontology of nature, acceptance of a specific set of rules about how to investigate nature, and adherence to a set of cognitive values about the teleology of natural inquiry (i.e., about the goals that science seeks). On this analysis, large-scale scientific change involves the replacement of one such world view by another, a process that entails the simultaneous repudiation of the key elements of the old picture and the adoption of corresponding (but of course different) elements of the new. In short, scientific change looks something like figure 1.

WV1 (ontology 1, methodology 1, values 1)



WV2 (ontology 2, methodology 2, values 2)

Fig. 1. Kuhn's Picture of Theory Change

When scientific change is construed so globally, it is no small challenge to see how it could be other than a conversion experience. If different scientists not only espouse different theories but also subscribe to different standards of appraisal and ground those standards in different and conflicting systems of cognitive goals, then it is difficult indeed to imagine that scientific change could be other than a whimsical change of style or taste. There could apparently never be compelling grounds for saying that one paradigm is better than another, for one has to ask: Better relative to which standards and whose goals? To make matters worse—much worse—Kuhn often suggested that each paradigm is more or less automatically guaranteed to satisfy its own standards and to fail the standards of rival paradigms, thus producing a kind of self-reinforcing solipsism in science. As he once put it, "To the extent, as significant as it is incomplete, that two scientific schools disagree about what is a problem and what a solution, they will inevitably talk through each other when debating the merits of their respective paradigms. In the partially circular arguments that regularly result, each paradigm will be shown to satisfy more or less the criteria that it dictates for itself and to fall short of a few of those dictated by its opponent."⁸ Anyone who writes prose of this sort must think that scientific decision making is fundamentally capricious. Or at least so many of us thought in the mid- and late 1960s, as philosophers began to digest Kuhn's ideas. In fact, if one looks at several discussions of Kuhn's work dating from that period, one sees this theme repeatedly. Paradigm change, it was said, could not possibly be a reasoned or rational process. Kuhn, we thought, has made science into an irrational "monster."

Kuhn's text added fuel to the fire by seeming to endorse such a construal of his own work. In a notorious discussion of the shift from the chemistry of Priestley to that of Lavoisier and Dalton, for instance, Kuhn asserted that it was perfectly reasonable for Priestley to hold onto phlogiston theory, just as it was fully rational for most of his contemporaries to be converting to the oxygen theory of Lavoisier. According to Kuhn, Priestley's continued adherence to phlogiston was reasonable because—given Priestley's cognitive aims and the methods he regarded as appropriate—his own theory continued to look good. Priestley lost the battle with Lavoisier, not because Priestley's paradigm was objectively inferior to its ri-

vals, but rather because most of the chemists of the day came to share Lavoisier's and Dalton's views about what was important and how it should be investigated.

The clear implication of such passages in Kuhn's writings is that interparadigmatic debate is necessarily inconclusive and thus can never be brought to rational closure. When closure does occur, it must therefore be imposed on the situation by such external factors as the demise of some of the participants or the manipulation of the levers of power and reward within the institutional structure of the scientific community. Philosophers of science, almost without exception, have found such implications troubling, for they directly confute what philosophers have been at pains for two millennia to establish: to wit, that scientific disputes, and more generally all disagreements about matters of fact, are in principle open to rational clarification and resolution. It is on the strength of passages such as those I have mentioned that Kuhn has been charged with relativism, subjectivism, irrationalism, and a host of other sins high on the philosopher's hit list.

There is some justice in these criticisms of Kuhn's work, for . . . Kuhn has failed over the past twenty years to elaborate any coherent account of consensus formation, that is, of the manner in which scientists could ever agree to support one world view rather than another. But that flaw, serious though it is, can probably be remedied, for I want to suggest that the problem of consensus formation can be solved if we will make two fundamental amendments in Kuhn's position. First . . . , we must replace the hierarchical view of justification with the reticulated picture, thereby making cognitive values "negotiable." Second, we must simply drop Kuhn's insistence on the integral character of world views or paradigms. More specifically, we solve the problem of consensus once we realize that *the various components of a world view are individually negotiable and individually replaceable in a piecemeal fashion* (that is, in such a manner that replacement of one element need not require wholesale repudiation of all the other components). Kuhn himself grants, of course, that some components of a world view can be revised; that is what "paradigm articulation" is all about. But for Kuhn, as for such other world view theorists as Lakatos and Foucault, the central commitments of a world view, its "hard core" (to use Lakatos's marvelous phrase), are not revisable—short of rejecting the entire world view. The core ontology of a world view or paradigm, along with its methodology and axiology, comes on a take-it-or-leave-it basis. Where these levels of commitment are concerned, Kuhn (along with such critics of his as Lakatos) is an uncompromising holist. Consider, for instance, his remark: "Just because it is a transition between incommensurables, the transition between competing paradigms cannot be made a step at a time . . . like the Gestalt-switch, it must occur all at once or not at all."⁹ Kuhn could hardly be less ambiguous on this point.

But paradigms or research programs need not be so rigidly conceived, and typically they are not so conceived by scientists; nor, if we reflect on it a moment, should they be so conceived. . . . There are complex justificatory interconnections among a scientist's ontology, his methodology, and his axiology. If a scientist's methodology fails to justify his ontology; if his methodology fails to promote his cognitive aims; if his cognitive aims prove to be utopian—in all these cases the scientist will have compelling reasons for replacing one component or other of his world view with an element that does the job better. Yet he need not modify everything else.

To be more precise, the choice confronting a scientist whose world view is under strain in this manner need be nothing like as stark as the choice sketched in figure 1 (where it is a matter of sticking with what he knows best unchanged or throwing that over for something completely different), but rather a choice where the modification of one core element—while retaining the others—may bring a decided improvement. Schematically, the choice may be one between

$$O_1 \text{ \& } M_1 \text{ \& } A_1 \quad (1)$$

and

$$O_2 \text{ \& } M_1 \text{ \& } A_1. \quad (2)$$

Or, between (1) and

$$O_1 \text{ \& } M_2 \text{ \& } A_1. \quad (3)$$

Or, to exhaust the simple cases, it may be between (1) and

$$O_1 \text{ \& } M_1 \text{ \& } A_2. \quad (4)$$

. . . Choices like those between (1) and (2), or between (1) and (3), are subject to strong normative constraints. And . . . choices of the sort represented between (1) and (4) are also, under certain circumstances, equally amenable to rational analysis.*

In all these examples there is enough common ground between the rivals to engender hope of finding an "Archimedean standpoint" which can rationally mediate the choice. When such commonality exists, there is no reason to regard the choice as just a matter of taste or whim; nor is there any reason to say of such choices, as Kuhn does (recall his characterization of the Priestley-Lavoisier exchange), that there can be no compelling grounds for one preference over another. Provided theory change

* Laudan defends these two claims in earlier chapters of *Science and Values*: the first in chapter 2, the second in chapter 3.

occurs one level at a time, there is ample scope for regarding it as a thoroughly reasoned process.

But the crucial question is whether change actually does occur in this manner. If one thinks quickly of the great transitions in the history of science, they *seem* to preclude such a stepwise analysis. The shift from (say) an Aristotelian to a Newtonian world view clearly involved changes on all three levels. So, too, did the emergence of psychoanalysis from nineteenth-century mechanistic psychology. But before we accept this wholesale picture of scientific change too quickly, we should ask whether it might not acquire what plausibility it enjoys only because our characterizations of such historical revolutions make us compress or telescope a number of gradual changes (one level at a time, as it were) into what, at our distance in time, can easily appear as an abrupt and monumental shift.

By way of laying out the core features of a more gradualist (and, I argue, historically more faithful) picture of scientific change, I will sketch a highly idealized version of theory change. Once it is in front of us, I will show in detail how it makes sense of some real cases of scientific change. Eventually, we will want a model that can show how one might move from an initial state of disagreement between rival traditions or paradigms to consensus about which one is better. But, for purposes of exposition, I want to begin with a rather simpler situation, namely, one in which consensus in favor of one world view or tradition gives way eventually to consensus in favor of another, without scientists ever being faced with a choice as stark as that between two well-developed, and totally divergent, rival paradigms. My "tall tale," represented schematically in figure 2, might go like this: at any given time, there will be at least one set of values, methods, and theories which one can identify as operating in any field or subfield of science. Let us call this collective C_1 , and its components, T_1 , M_1 , and A_1 . These components typically stand in . . . complex justificatory relationships to one another . . . ; that is, A_1 will justify M_1 and harmonize with T_1 ; M_1 will justify T_1 and exhibit the realizability of A_1 ; and T_1 will constrain M_1 and exemplify A_1 . Let us suppose that someone then proposes a new theory, T_2 , to replace T_1 . The rules M_1 will be consulted and they may well indicate grounds for preferring T_2 to T_1 . Suppose that they do, and that we thereby replace T_1 with T_2 . As time goes by, certain scientists may develop reservations about M_1 and propose a new and arguably superior methodology, M_2 . Now a choice must be made between M_1 and M_2 . As we have seen, that requires determining whether M_1 or M_2 offers more promise of realizing our aims. Since that determination will typically be an empirical matter, both A_1 and the then prevailing theory, T_2 , will have to be consulted to ascertain whether M_1 or M_2 is optimal for securing A_1 . Suppose that, in comparing the relative efficacy of achieving the shared values, A_1 , cogent arguments can be made to show that M_2 is superior to M_1 . Under the circumstances, assuming

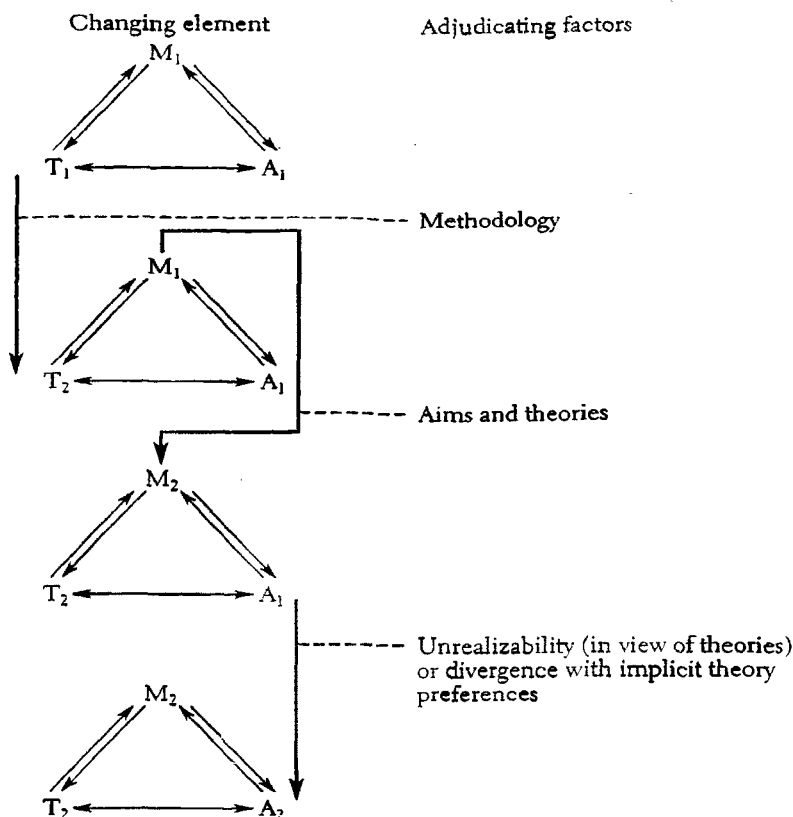


Fig. 2. Untraditional Change

scientists behave rationally, M_2 will replace M_1 . This means that as new theories, T_3, T_4, \dots, T_n , emerge later, they will be assessed by rules M_2 rather than M_1 . Suppose, still further along in this fairy tale, we imagine a challenge to the basic values themselves. Someone may, for instance, point to new evidence suggesting that some element or other of A_1 is unrealizable. Someone else may point out that virtually none of the theories accepted by the scientific community as instances of good science exemplify the values expressed in A_1 . (Or, it may be shown that A_1 is an inconsistent set in that its component aspirations are fundamentally at odds with one another.) Under such circumstances, scientists may rationally decide to abandon A_1 and to take up an alternative, consistent set of values,

A_2 , should it be available. (Although I have considered a temporal sequence of changes—first in theory, then in methods, and finally in aims—which superficially corresponds to the justificatory order of the hierarchical model, it is crucial to realize how unlike the hierarchical picture this sequence really is. That model would countenance no rational deliberation of the sort represented by the transition from $T_2M_2A_1$ to $T_2M_2A_2$. Equally, the hierarchical model, . . . does not permit our beliefs at the level of theories to shape our views as to permissible methods, since justification in the hierarchical model is entirely downward from methods to theories.)

Now that we have this hypothetical sequence before us, let us imagine a historian called Tom, who decides many years later to study this episode. He will doubtless be struck by the fact that a group of scientists who once accepted values A_1 , rules M_1 , and theory T_1 came over the course of, say, a decade or two to abandon the whole lot and to accept a new complex, C_2 , consisting of A_2 , M_2 , and T_2 . Tom will probably note, and rightly too, that the partisans of C_2 have precious little in common with the devotees of C_1 . Surely, Tom may well surmise, here was a scientific revolution if ever there was one, for there was dramatic change at every level. If Tom decides to call the view that scientists eventually came to hold "Paradigm 2," and the view from which they began "Paradigm 1," then he will be able to document the existence of a massive paradigm shift between what (at our remoteness in time) appear to be conceptually distant and virtually incommensurable paradigms.

The point, of course, is that a sequence of belief changes which, described at the microlevel, appears to be a perfectly reasonable and rational sequence of events may appear, when represented in broad brushstrokes that drastically compress the temporal dimension, as a fundamental and unintelligible change of world view. This kind of tunnel vision, in which a sequence of gradual shifts is telescoped into one abrupt and mighty transformation, is a folly which every historian is taught to avoid. Yet knowing that one should avoid it and actually doing so are two different things. Once we recognize this fallacy for what it is, we should probably hesitate to accept too quickly the models of the holists and big-picture builders. For, if our fairy story has anything of the truth about it (that is, if change is, or more weakly even if it could be, more piecemeal than the holistic accounts imply), there may yet be room for incorporating changes of methods and of cognitive values into a rational account of scientific activity. My object in the rest of this chapter is to offer some reasons to believe that the fairy tale is a good deal closer to the mark than its holistic rivals.

But before I present the evidence needed for demythologizing my story, we have to add a new twist to it. As I pointed out above, this story concerns what I call a "unitraditional paradigm shift." It reveals how it might be possible for scientists, originally advocates of one tradition or

paradigm, to come around eventually to accept what appears to be a very different view of the world, not to say a very different view of what science is. I call such a change untraditional because it is not prompted or provoked by the availability of a well-articulated rival world view. If you like, the untraditional picture explains how one could get paradigm change by developments entirely internal to the dynamic of a particular paradigm. More interesting, and more challenging, is the problem of multitransitional paradigm shifts, that is, basic changes of world view which arise from competition between rival paradigms. To deal with such cases, we need to complicate our fairy story a bit.

Here, we need to imagine two of our complexes already well developed, and radically divergent (i.e., with different ontologies, different methodologies, and different axiologies). If we ask under what circumstances it would be reasonable for the partisans of C_1 to abandon it and accept C_2 , some answers come immediately to mind. Suppose, for instance, it can be shown that the central theories of C_1 look worse than the theories of C_2 , even by the standards of C_1 . As we have seen, Kuhn denies that this is possible, since he says that the theories associated with a particular paradigm will always look better by its standards than will the theories of rival paradigms.¹⁰ But as we have already seen, there is no way of guaranteeing in advance that the methods and standards of C_1 will always give the epistemic nod to theories associated with C_1 , since it is always possible (and has sometimes happened) that rival paradigms to C_1 will develop theories that do a better job of satisfying the methodological demands of C_1 than do the theories developed within C_1 itself. Alternatively, suppose someone shows that there is a set of methods M_2 which is more nearly optimal than M_1 for achieving the aims of C_1 , and that those methods give the epistemic nod to the theories of C_2 rather than those of C_1 . Or, suppose that someone shows that the goals of C_1 are deeply at odds with the attributes of some of the major theories of science—theories that the partisans of C_1 themselves endorse—and that, by contrast, the cognitive values of C_2 are typified by those same theories. Again, new evidence might emerge which indicates the nonrealizability of some of the central cognitive aims of C_1 and the achievability of the aims of C_2 . In all these circumstances (and several obvious ones which I shall not enumerate), the only reasonable thing for a scientist to do would be to give up C_1 and to embrace C_2 .

But, once we begin to play around with the transformations permitted by the reticulation model, we see that the transition from one paradigm or world view to another can itself be a step-wise process, requiring none of the wholesale shifts in allegiance at every level required by Kuhn's analysis. The advocates of C_1 might, for instance, decide initially to accept many of the substantive theories of C_2 , while still retaining for a time the methodology and axiology of C_1 . At a later stage they might be led by a different chain of arguments and evidence to accept the methodology of

C_2 while retaining C_1 's axiology. Finally, they might eventually come to share the values of C_2 . As William Whewell showed more than a century ago, precisely some such series of shifts occurred in the gradual capitulation of Cartesian physicists to the natural philosophy of Newton.¹¹

In effect, I am claiming that the solution of the problem of consensus formation in the multiparadigm situation to be nothing more than a special or degenerate instance of untraditional change. It follows that, if we can show that the untraditional fairy tale has something going for it, then we will solve both forms of the consensus-formation problem simultaneously. The core question is whether the gradualist myth, which I have just sketched out, is better supported by the historical record than the holistic picture associated with Kuhn.

One striking way of formulating the contrast between the piecemeal and the holistic models, and thus designing a test to choose between them, is to ask a fairly straightforward question about the historical record: Is it true that the major historical shifts in the methodological rules of science and in the cognitive values of scientists have invariably been contemporaneous with one another *and* with shifts in substantive theories and ontologies? The holistic account is clearly committed to an affirmative answer to the question. Indeed, it is a straightforward corollary of Kuhn's analysis that changes in rules or values, when they occur, will occur only when a scientific revolution takes place, that is, only when there is a concomitant shift in theories, methods, and values. A change in values without an associated change in basic ontology is not a permissible variation countenanced in the Kuhnian scheme.¹² Nor is a change in methods possible for Kuhn without a paradigm change. Kuhn's analysis flatly denies that the values and norms of a "mature" science can shift in the absence of a revolution. Yet there are plenty of examples one may cite to justify the assertion made here that changes at the three levels do not always go together. I shall mention two such examples.

Consider, first, a well-known shift at the level of methodological rules. From the time of Bacon until the early nineteenth century most scientists subscribed to variants of the rules of inductive inference associated with Bacon, Hume, and Newton. The methods of agreement, difference, and concomitant variations were a standard part of the repertoire of most working scientists for two hundred years. These rules, at least as then understood, foreclosed the postulation of any theoretical or hypothetical entities, since observable bodies were the only sort of objects and properties to which one could apply traditional inductive methods. More generally . . . , thinkers of the Enlightenment believed it important to develop rules of inquiry which would exclude unobservable entities and bring to heel the tendency of scientists to indulge their *esprit de système*. Newton's famous third rule of reasoning in philosophy, the notorious "hypotheses

non fingo," was but a particularly succinct and influential formulation of this trenchant empiricism.*

It is now common knowledge that by the late nineteenth century this methodological orientation had largely vanished from the writings of major scientists and methodologists. Whewell, Peirce, Helmholtz, Mach, Darwin, Hertz, and a host of other luminaries had, by the 1860s and 1870s, come to believe that it was quite legitimate for science to postulate unobservable entities, and that most of the traditional rules of inductive reasoning had been superseded by the logic of hypothetico-deduction. Elsewhere I have described this shift in detail.¹³ What is important for our purposes is both that it occurred and when it occurred. That it took place would be denied, I think, by no one who studies the record; determining precisely when it occurred is more problematic, although probably no scholar would quarrel with the claim that it comes in the period from 1800 to 1860. And a dating as fuzzy as that is sufficient to make out my argument.

For here we have a shift in the history of the explicit methodology of the scientific community as significant as one can imagine—from methods of enumerative and eliminative induction to the method of hypothesis—occurring across the spectrum of the theoretical sciences, from celestial mechanics to chemistry and biology.¹⁴ Yet where is the larger and more global scientific revolution of which this methodological shift was the concomitant? There were of course revolutions, and important ones, in this period. Yet this change in methodology cannot be specifically linked to any of the familiar revolutions of the period. The method of hypothesis did not become the orthodoxy in science of the late nineteenth century because it rode on the coattails of any specific change in ontology or scientific values. So far as I can see, this methodological revolution was independent of any particular program of research in any one of the sciences, which is not to say that it did not reflect some very general tendencies appearing across the board in scientific research. The holist

* Newton's phrase *hypotheses non fingo* (I feign no hypotheses) occurs in the General Scholium of the *Principia*, not in the third rule of reasoning in philosophy. Having discovered the laws according to which gravity operates, Newton declares that he does not intend to hypothesize about the cause of gravitational attraction. Newton writes: "But hitherto I have not been able to discover the cause of those properties of gravity from phenomena, and I feign no hypotheses; for whatever is not deduced from the phenomena is to be called an hypothesis; and hypotheses, whether metaphysical or physical, whether of occult qualities or mechanical, have no place in experimental philosophy." This widely quoted and oft-criticized manifesto of inductive empiricism is an elaboration of Newton's fourth rule of reasoning in philosophy. See the selection from Duhem in chapter 3 of this volume for a criticism of Newton's claim that the law of universal gravitation can be "deduced from the phenomena." Isaac Newton, *Philosophiæ naturalis principia mathematica*, vol. 2, trans. A. Motte, rev. F. Cajori (Berkeley: University of California Press, 1934), 547.

model, which would have us believe that changes in methodological orientation are invariably linked to changes in values and ontology, is patently mistaken here. Nor, if one reflects on the nature of methodological discussion, should we have expected otherwise. . . . Methodological rules can reasonably be criticized and altered if one discovers that they fail optimally to promote our cognitive aims. If our aims shift, as they would in a Kuhnian paradigm shift, we would of course expect a reappraisal of our methods of inquiry in light of their suitability for promoting the new goals. But, even when our goals shift not at all, we sometimes discover arguments and evidence which indicate that the methods we have been using all along are not really suitable for our purposes. Such readjustments of methodological orientation, in the absence of a paradigm shift, are a direct corollary of the reticulational model as I described it earlier; yet they pose a serious anomaly for Kuhn's analysis.

What about changes in aims, as opposed to rules? Is it not perhaps more plausible to imagine, with Kuhn, that changes of cognitive values are always part of broader shifts of paradigm or world view? Here again, the historical record speaks out convincingly against this account. Consider, very briefly, one example: the abandonment of "infallible knowledge" as an epistemic aim for science. As before, my historical account will have to be "potted" for purposes of brevity; but there is ample serious scholarship to back up the claims I shall be making.¹⁵

That scholarship has established quite convincingly that, during the course of the nineteenth century, the view of science as aiming at certainty gave way among most scientists to a more modest program of producing theories that were plausible, probable, or well tested. As Peirce and Dewey have argued, this shift represents one of the great watersheds in the history of scientific philosophy: the abandonment of the quest for certainty. More or less from the time of Aristotle onward, scientists had sought theories that were demonstrable and apodictically certain. Although empiricists and rationalists disagreed about precisely how to certify knowledge as certain and incorrigible, all agreed that science was aiming exclusively at the production of such knowledge. This same view of science largely prevailed at the beginning of the nineteenth century. But by the end of that century this demonstrative and infallibilist ideal was well and truly dead. Scientists of almost every persuasion were insistent that science could, at most, aspire to the status of highly probable knowledge. Certainty, incorrigibility, and indefeasibility ceased to figure among the central aims of most twentieth-century scientists.

The full story surrounding the replacement of the quest for certainty by a thoroughgoing fallibilism is long and complicated; I have attempted to sketch out parts of that story elsewhere.¹⁶ What matters for our purposes here is not so much the details of this epistemic revolution, but the fact that this profound transformation was not specifically associated with the emergence of any new scientific paradigms or research programs. The

question of timing is crucial, for it is important to see that this deep shift in axiological sensibilities was independent of any specific change in scientific world view or paradigm. No new scientific tradition or paradigm in the nineteenth century was associated with a specifically fallibilist axiology. Quite the reverse, fallibilism came to be associated with virtually every major program of scientific research by the mid- to late nineteenth century. Atomists and antiatomists, wave theorists and particle theorists, Darwinians and Lamarckians, uniformitarians and catastrophists—all subscribed to the new consensus about the corrigibility and indemonstrability of scientific theories. A similar story could be told about other cognitive values which have gone the way of all flesh. The abandonment of intelligibility, of the requirement of picturable or mechanically constructible models of natural processes, of the insistence on “complete” descriptions of nature—all reveal a similar pattern. The abandonment of each of these cognitive ideals was largely independent of shifts in basic theories about nature.

Once again, the holistic approach leads to expectations that are confounded by the historical record. Changes in values and changes in substantive ontologies or methodologies show no neat isomorphism. Change certainly occurs at all levels, and sometimes changes are concurrent, but there is no striking covariance between the timing of changes at one level and the timing of those at any other. I conclude from such examples that scientific change is substantially more piecemeal than the holistic model would suggest. Value changes do not always accompany, nor are they always accompanied by, changes in scientific paradigm. Shifts in methodological rules may, but need not, be associated with shifts in either values or ontologies. The three levels, although unquestionably interrelated, do not come as an inseparable package on a take-it-or-leave-it basis.

This result is of absolutely decisive importance for understanding the processes of scientific change. Because these changes are not always concomitant, we are often in a position to hold one or two of the three levels fixed while we decide whether to make modifications at the disputed level. The existence of these (temporarily) fixed and thus shared points of perspective provides a crucial form of triangulation. Since theories, methodologies, and axiologies stand together in a kind of justificatory triad, we can use those doctrines about which there is agreement to resolve the remaining areas where we disagree. The uncontested levels will not always resolve the controversy, for underdetermination is an ever present possibility. But the fact that the levels of agreement are sometimes insufficient to terminate the controversy provides no comfort for Kuhn's subjectivist thesis that those levels of agreement are never sufficient to resolve the debate. As logicians say, we need to be very careful about our quantifiers here. Some writers have not always exercised the care they should. Kuhn, for instance, confusedly slides from (a) the correct claim that the shared values of scientists are, in certain situations, incapable of yielding unam-

biguously a preference between two rival theories to (b) the surely mistaken claim that the shared values of scientists are never sufficient to warrant a preference between rival paradigms. Manifestly in some instances, the shared rules and standards of methodology are unavailing. But neither Kuhn nor anyone else has established that the rules, evaluative criteria, and values to which scientists subscribe are generally so ambiguous in application that virtually any theory or paradigm can be shown to satisfy them. And we must constantly bear in mind the point that, even when theories are underdetermined by a set of rules or standards, many theories will typically be ruled out by the relevant rules; and if one party to a scientific debate happens to be pushing for a theory that can be shown to violate those rules, then the rules will eliminate that theory from contention.

What has led holistic theorists to misdescribe so badly the relations among these various sorts of changes? As one who was himself once an advocate of such an account, I can explain specifically what led me into thinking that change on the various levels was virtually simultaneous. If one focuses, as most philosophers of science have, on the processes of justification in science, one begins to see systemic linkages among what I earlier called *factual*, *methodological*, and *axiological* levels. One notices further that beliefs at all three levels shift through time. Under the circumstances it is quite natural to conjecture that these various changes may be interconnected. Specifically, one can imagine that the changes might well be simultaneous, or at least closely dependent on one another. The suggestion is further borne out—at least to a first approximation—by an analysis of some familiar scientific episodes. It is clear, for instance, that the scientific revolution of the seventeenth century brought with it changes in theories, ontologies, rules, and values. Equally, the twentieth-century revolution in relativity theory and quantum mechanics brought in its wake a shift in both methodological and axiological orientations among theoretical physicists. But as I have already suggested, these changes came *seriatim*, not simultaneously. More to the point, it is my impression that the overwhelming majority of theory transitions in the history of science (including shifts as profound as that from creationist biology to evolution, from energeticist to atomistic views on the nature of matter, from catastrophism to uniformitarianism in geology, from particle to wave theories of light) have not taken place by means of Gestalt-like shifts at all levels concurrently. Often, change occurs on a single level only (e.g., the Darwinian revolution or the triumph of atomism, where it was chiefly theory or ontology that changed); sometimes it occurs on two levels simultaneously; rarely do we find an abrupt and wholesale shift of doctrines at all three levels.

This fact about scientific change has a range of important implications for our understanding of scientific debate and scientific controversy. Leaving aside the atypical case of simultaneous shifts at all three levels . . . , it

means that most instances of scientific change—including most of the events we call scientific revolutions—occur amid a significant degree of consensus at a variety of levels among the contending parties. Scientists may, for instance, disagree about specific theories yet agree about the appropriate rules for theory appraisal. They may even disagree about both theories and rules but accept the same cognitive values. Alternatively, they may accept the same theories and rules yet disagree about the cognitive values they espouse. In all these cases there is no reason to speak (with Kuhn) of “incommensurable choices” or “conversion experiences,” or (with Foucault) about abrupt “ruptures of thought,” for there is in each instance the possibility of bringing the disagreement to rational closure. Of course, it may happen in specific cases that the mechanisms of rational adjudication are of no avail, for the parties may be contending about matters that are underdetermined by the beliefs and standards the contending parties share in common. But, even here, we can still say that there are rational rules governing the game being played, and that the moves being made (i.e., the beliefs being debated and the arguments being arrayed for and against them) are in full compliance with the rules of the game.

Above all, we must bear in mind that it has never been established that such instances of holistic change constitute more than a tiny fraction of scientific disagreements. Because such cases are arguably so atypical, it follows that sociologists and philosophers of science who predicate their theories of scientific change and cognition on the presumed ubiquity of irresolvable standoffs between monolithic world views (of the sort that Kuhn describes in *The Structure of Scientific Revolutions*) run the clear risk of failing to recognize the complex ways in which rival theories typically share important background assumptions in common. To put it differently, global claims about the immunity of interparadigmatic disputes to rational adjudication (and such claims are central in the work of both Kuhn and Lakatos) depend for their plausibility on systematically ignoring the piecemeal character of most forms of scientific change and on a gross exaggeration of the impotence of rational considerations to bring such disagreements to closure. Beyond that, I have argued that, even if interparadigmatic clashes had the character Kuhn says they do (namely, of involving little or no overlap at any of the three levels), it still would not follow that there are no rational grounds for a critical and comparative assessment of the rival paradigms. In sum, no adequate support has been provided for the claim that clashes between rival scientific camps can never, or rarely ever, be resolved in an objective fashion. The problem of consensus formation, which . . . was the great Kuhnian enigma,¹⁷ can be resolved, but only if we realize that science has adjudicatory mechanisms whose existence has gone unnoticed by Kuhn and the other holists.

But it would be misleading to conclude this treatment of Kuhn and the holist theory of theory change on such a triumphal note, for we have

yet to confront directly and explicitly another relevant side of Kuhn's work: specifically, his claim, elaborated through a variety of arguments, that methodological rules and shared cognitive values (on which I have laid so much stress as instruments of closure and consensus formation) are impotent to resolve large-scale scientific disagreement. We must now turn to that task directly.

■ | Kuhn's Critique of Methodology

Several writers (e.g., Quine, Hesse, Goodman) have asserted that the rules or principles of scientific appraisal underdetermine theory choice. For reasons I have tried to spell out elsewhere,¹⁸ such a view is badly flawed. Some authors, for instance, tend to confuse the logical underdetermination of theories by data with the underdetermination of theory choice by methodological rules. Others (e.g., Hesse and Bloor) have mistakenly taken the logical underdetermination of theories to be a license for asserting the causal underdetermination of our theoretical beliefs by the sensory evidence to which we are exposed.¹⁹ But there is a weaker, and much more interesting, version of the thesis of underdetermination, which has been developed most fully in Kuhn's recent writings. Indeed, it is one of the strengths of Kuhn's challenge to traditional philosophy of science that he has "localized" and given flesh to the case for underdetermination, in ways that make it *prima facie* much more telling. In brief, Kuhn's view is this: if we examine situations where scientists are required to make a choice among the handful of paradigms that confront them at any time, we discover that the relevant evidence and appropriate methodological standards fail to pick out one contender as unequivocally superior to its extant rival(s). I call such situations cases of "local" underdetermination, by way of contrasting them with the more global forms of underdetermination (which say, in effect, that the rules are insufficient to pick out any theory as being uniquely supported by the data). Kuhn offers four distinct arguments for local underdetermination. Each is designed to show that, although methodological rules and standards do constrain and delimit a scientist's choices or options, those rules and standards are never sufficient to compel or unequivocally to warrant the choice of one paradigm over another.

I THE "AMBIGUITY OF SHARED STANDARDS" ARGUMENT

Kuhn's first argument for methodological underdetermination rests on the purported ambiguity of the methodological rules or standards that are shared by advocates of rival paradigms. The argument first appeared in *The Structure of Scientific Revolutions* (1962) and has been extended con-

siderably in his later *The Essential Tension* (1977). As he put it in the earlier work, "lifelong resistance [to a new theory] . . . is not a violation of scientific standards . . . though the historian can always find men—Priestley, for instance—who were unreasonable to resist for as long as they did, he will not find a point at which resistance becomes illogical or unscientific."²⁰ Many of Kuhn's readers were perplexed by the juxtaposition of claims in such passages as these. On the one hand, we are told that Priestley's continued refusal to accept the theory of Lavoisier was "unreasonable"; but we are also told that Priestley's refusal was neither "illogical" nor "unscientific." To those inclined to think that being "scientific" (at least in the usual sense of that term) required one to be "reasonable" about shaping one's beliefs, Kuhn seemed to be talking gibberish. On a more sympathetic construal, Kuhn seemed to be saying that a scientist could always interpret the applicable standards of appraisal, whatever they might be, so as to "rationalize" his own paradigmatic preferences, whatever they might be. This amounts to claiming that the methodological rules or standards of science never make a real or decisive difference to the outcome of a process of theory choice; if any set of rules can be used to justify any theory whatever, then methodology would seem to amount to just so much window dressing. But that construal, it turns out, is a far cry from what Kuhn intended. As he has made clear in later writings, he wants to bestow a positive, if (compared with the traditional view) much curtailed, role on methodological standards in scientific choice.

What Kuhn apparently has in mind is that the shared criteria, standards, and rules to which scientists explicitly and publicly refer in justifying their choices of theory and paradigm are typically "ambiguous" and "imprecise," so much so that "individuals [who share the same standards] may legitimately differ about their application to concrete cases."²¹ Kuhn holds that, although scientists share certain cognitive values "and must do so if science is to survive, they do not all apply them in the same way. Simplicity, scope, fruitfulness, and even accuracy can be judged differently (which is not to say they may be judged arbitrarily) by different people."²² Because, then, the shared standards are ambiguous, two scientists may subscribe to "exactly the same standard" (say, the rule of simplicity) and yet endorse opposing viewpoints.

Kuhn draws some quite large inferences from the presumed ambiguity of the shared standards or criteria. Specifically, he concludes that every case of theory choice must involve an admixture of objective and subjective factors, since (in Kuhn's view) the shared, and presumably objective, criteria are too amorphous and ambiguous to warrant a particular preference. He puts the point this way: "I continue to hold that the algorithms of individuals are all ultimately different by virtue of the subjective considerations with which each [scientist] must complete the objective criteria before any computations can be done."²³ As this passage makes clear, Kuhn believes that, because the shared criteria are too imprecise to justify

a choice, and because—despite that imprecision—scientists do manage to make choices, those choices *must* be grounded in individual and subjective preferences different from those of his fellow scientists. As he says, “every individual choice between competing theories depends on a mixture of objective and subjective factors, or of shared and individual criteria.”²⁴ And, the shared criteria “are not by themselves sufficient to determine the decisions of individual scientists.”²⁵

This very ambitious claim, if true, would force us to drastically rethink our views of scientific rationality. Among other things, it would drive us to the conclusion that every scientist has different reasons for his theory preferences from those of his fellow scientists. The view entails, among other things, that it is a category mistake to ask (say) why physicists think Einstein’s theories are better than Newton’s; for, on Kuhn’s analysis, there must be as many different answers as there are physicists. We might note in passing that this is quite an ironic conclusion for Kuhn to reach. Far more than most writers on these subjects, he has tended to stress the importance of community and socialization processes in understanding the scientific enterprise. Yet the logic of his own analysis drives him to the radically individualistic position that every scientist has his own set of reasons for theory preferences and that there is no real consensus whatever with respect to the grounds for theory preference, not even among the advocates of the same paradigm. Seen from this perspective, Kuhn tackles what I earlier called the problem of consensus by a maneuver that trivializes the problem: for if we must give a separate and discrete explanation for the theory preferences of each member of the scientific community—which is what Kuhn’s view entails—then we are confronted with a gigantic mystery at the collective level, to wit, why the scientists in a given discipline—each supposedly operating with his own individualistic and idiosyncratic criteria, each giving a different “gloss” to the criteria that are shared—are so often able to agree about which theories to bet on. But we can leave it to Kuhn to sort out how he reconciles his commitment to the social psychology of science with his views about the individual vagaries of theory preference. What must concern us is the question whether Kuhn has made a plausible case for thinking that the shared or collective criteria must be supplemented by individual and subjective criteria.

The first point to stress is that Kuhn’s thesis purports to apply to all scientific rules or values that are shared by the partisans of rival paradigms, not just to a selected few, notoriously ambiguous ones. We can grant straightaway that some of the rules, standards, and values used by scientists (“simplicity” would be an obvious candidate) exhibit precisely that high degree of ambiguity which Kuhn ascribes to them. But Kuhn’s general argument for the impotence of shared rules to settle disagreements between scientists working in different paradigms cannot be established by citing the occasional example. Kuhn must show us, for he claims as much, that there is something in the very nature of those methodological rules

that come to be shared among scientists which makes the application of those rules or standards invariably inconclusive. He has not established this result, and there is a good reason why he has not: it is false. To see that it is, one need only produce a methodological rule widely accepted by scientists which can be applied to concrete cases without substantial imprecision or ambiguity. Consider, for instance, one of Kuhn's own examples of a widely shared scientific standard, namely, the requirement that an acceptable theory must be internally consistent and logically consistent with accepted theories in other fields. (One may or may not favor this methodological rule. I refer to it here only because it is commonly regarded, including by Kuhn, as a methodological rule that frequently plays a role in theory evaluation.)

I submit that we have a very clear notion of what it is for a theory to be internally consistent, just as we understand perfectly well what it means for a theory to be consistent with accepted beliefs. Moreover, on at least some occasions we can tell whether a particular theory has violated the standard of (internal or external) consistency. Kuhn himself, in a revealing passage, grants as much; for instance, when comparing the relative merits of geocentric and heliocentric astronomy, Kuhn says that "the consistency criterion, by itself, therefore, spoke unequivocally for the geocentric tradition."²⁶ (What he has in mind is the fact that heliocentric astronomy, when introduced, was inconsistent with the then reigning terrestrial physics, whereas the assumptions of geocentric astronomy were consistent with that physics.) Note that in this case we have a scientific rule or criterion "speaking unequivocally" in favor of one theory and against its rival. Where are the inevitable imprecision and ambiguity which are supposed by Kuhn to afflict all the shared values of the scientific community? What is ambiguous about the notion of consistency? The point of these rhetorical questions is to drive home the fact that, even by Kuhn's lights, some of the rules or criteria widely accepted in the scientific community do not exhibit that multiplicity of meanings which Kuhn has described as being entirely characteristic of methodological standards.

One could, incidentally, cite several other examples of reasonably clear and unambiguous methodological rules. For instance, the requirements that theories should be deductively closed or that theories should be subjected to controlled experiments have not generated a great deal of confusion or disagreement among scientists about what does and does not constitute closure or a control. Or, consider the rule that theories should lead successfully to the prediction of results unknown to their discoverer; so far as I am aware, scientists have not differed widely in their construal of the meaning of this rule. The significance of the nonambiguity of many methodological concepts and rules is to be found in the fact that such nonambiguity refutes one of Kuhn's central arguments for the incomparability of paradigms and for its corollary, the impotence of methodology as a guide to scientific rationality. There are at least some rules that are

sufficiently determinate that one can show that many theories clearly fail to satisfy them. We need not supplement the shared content of these objective concepts with any private notions of our own in order to decide whether a theory satisfies them.

2 THE "COLLECTIVE INCONSISTENCY OF RULES" ARGUMENT

As if the ambiguity of standards was not bad enough, Kuhn goes on to argue that the shared rules and standards, when taken as a collective, "repeatedly prove to conflict with one another."²⁷ For instance, two scientists may each believe that empirical accuracy and generality are desirable traits in a theory. But, when confronted with a pair of rival (and thus incompatible) theories, one of which is more accurate and the other more general, the judgments of those scientists may well differ about which theory to accept. One scientist may opt for the more general theory; the other, for the more accurate. They evidently share the same standards, says Kuhn, but they end up with conflicting appraisals. Kuhn puts it this way: "... in many concrete situations, different values, though all constitutive of good reasons, dictate different conclusions, different choices. In such cases of value-conflict (e.g., one theory is simpler but the other is more accurate) the relative weight placed on different values by different individuals can play a decisive role in individual choice."²⁸

Because many methodological standards do pull in different directions, Kuhn thinks that the scientist can pretty well go whichever way he likes. Well, not quite any direction he likes, since—even by Kuhn's very liberal rules—it would be unreasonable for a scientist to prefer a theory (or paradigm) which failed to satisfy any of the constraints. In Kuhn's view, we should expect scientific disagreement or dissensus to emerge specifically in those cases where (a) no available theory satisfied all the constraints and (b) every extant theory satisfied some constraints not satisfied by its rivals. That scientists sometimes find themselves subscribing to contrary standards, I would be the first to grant. Indeed, ... the discovery of that fact about oneself is often the first prod toward readjusting one's cognitive values. But Kuhn is not merely saying that this happens occasionally; he is asserting that such is the nature of any set of rules or standards which any group of reasonable scientists might accept. As before, our verdict has to be that Kuhn's highly ambitious claim is just that; he never shows us why families of methodological rules should always or even usually be internally inconsistent. He apparently expects us to take his word for it that he is just telling it as it is.²⁹ I see no reason why we should follow Kuhn in his global extrapolations from the tiny handful of cases he describes. On the contrary, there are good grounds for resisting, since there are plenty of sets of consistent methodological standards. Consider, for instance, one of the most influential documents of nineteenth-century scientific methodology, John Stuart Mill's *System of Logic*. Mill offered

there a set of rules or canons for assessing the soundness of causal hypotheses. Nowadays those rules are still called "Mill's methods," and much research in the natural and social sciences utilizes them, often referring to them as the methods of agreement, difference, and concomitant variations. To the best of my knowledge, no one has ever shown that Mill's methods exhibit a latent tendency toward contradiction or conflict of the sort that Kuhn regards as typical of systems of methodological rules. To go back further in history, no one has ever shown that Bacon's or Descartes's or Newton's or Herschel's famous canons of reasoning are internally inconsistent. The fact that numerous methodologies of science may be cited which have never been shown to be inconsistent casts serious doubts on Kuhn's claim that any methodological standards apt to be shared by rival scientists will tend to exhibit mutual inconsistencies.

Kuhn could have strengthened his argument considerably if, instead of focusing on the purported tensions in sets of methodological rules, he had noted, rather, that whenever one has more than one standard in operation, it is conceivable that we will be torn in several directions. And this claim is true, regardless of whether the standards are strictly inconsistent with one another or not (just so long as there is not a complete covariance between their instances). If two scientists agree to judge theories by two standards, then it is trivially true that, depending upon how much weight each gives to the two standards, their judgments about theories may differ. Before we can make sense of how to work with several concurrent standards, we have to ask (as Kuhn never did) about the way in which these standards do (or should) control the selection of a preferred theory. Until we know the answer to that question, we will inevitably find that the standards are of little use in explaining scientific preferences. Kuhn simply assumes that all possible preference structures (i.e., all possible differential weightings of the applicable standards) are equally viable or equally likely to be exemplified in a working scientist's selection procedures. . . .

To sum up the argument to this point: I have shown that Kuhn is wrong in claiming that all methodological rules are inevitably ambiguous and in claiming that scientific methodologies consisting of whole groups of rules always or even usually exhibit a high degree of internal "tension." Since these two claims were the linchpins in Kuhn's argument to the effect that shared criteria "are not by themselves sufficient to determine the decisions of individual scientists,"³⁰ we are entitled to say that Kuhn's effort to establish a general form of local underdetermination falls flat.

3 THE SHIFTING STANDARDS ARGUMENT

Equally important to Kuhn's critique of methodology is a set of arguments having to do with the manner in which standards are supposed to vary from one scientist to another. In treating Kuhn's views on this matter, I

follow Gerald Doppelt's excellent and sympathetic explication of Kuhn's position.³¹ In general, Kuhn's model of science envisages two quite distinct ways in which disagreements about standards might render scientific debate indeterminate or inconclusive. In the first place, the advocates of different paradigms may subscribe to different methodological rules or evaluative criteria. Indeed, "may" is too weak a term here, for, as we have seen, Kuhn evidently believes that associated with each paradigm is a set of methodological orientations that are (at least partly) at odds with the methodologies of all rival paradigms. Thus, he insists that whenever a "paradigm shift" occurs, this process produces "changes in the standards governing permissible problems, concepts and explanations."³² This is quite a strong claim. It implies, among other things, that the advocates of different paradigms invariably have different views about what constitutes a scientific explanation and even about what constitutes the relevant facts to be explained (viz., the "permissible problems"). If Kuhn is right about these matters, then debate between the proponents of two rival paradigms will involve appeal to different sets of rules and standards associated respectively with the two paradigms. One party to the dispute may be able to show that his theory is best by his standards, while his opponent may be able to claim superiority by his.

As I have shown in detail earlier in this chapter, Kuhn is right to say that scientists sometimes subscribe to different methodologies (including different standards for explanation and facticity). But he has never shown, and I believe him to be chronically wrong in claiming, that disagreements about matters of standards and rules neatly coincide with disagreements about substantive matters of scientific ontology. Rival scientists advocating fundamentally different theories or paradigms often have the same standards of assessment (and interpret them identically); on the other hand, adherents to the same paradigm will frequently espouse different standards. In short, methodological disagreements and factual disagreements about basic theories show no striking covariances of the kind required to sustain Kuhn's argument about the intrinsic irresolvability of interparadigmatic debate. It was the thrust of my earlier account of "piecemeal change" to show why Kuhn's claims about irresolvability will not work.

But, of course, a serious issue raised by Kuhn still remains before us. If different scientists sometimes subscribe to different standards of appraisal (and that much is surely correct), then how is it possible for us to speak of the resolution of such disagreements as anything other than an arbitrary closure? To raise that question presupposes a picture of science which I [have] sought to demolish. . . . Provided there are mechanisms for rationally resolving disagreements about methodological rules and cognitive values . . . , the fact that scientists often disagree about such rules and values need not, indeed should not, be taken to show that there must be anything arbitrary about the resolution of such disagreements.

4 THE PROBLEM-WEIGHTING ARGUMENT

As I have said earlier, Kuhn has another argument up his sleeve which he and others think is germane to the issue of the rationality of comparative theory assessment. Specifically, he insists that the advocates of rival paradigms assign differential degrees of importance to the solution of different sorts of problems. Because they do, he says that they will often disagree about which theory is better supported, since one side will argue that it is most important to solve a certain problem, while the other will insist on the centrality of solving a different problem. Kuhn poses the difficulty in these terms: "if there were but one set of scientific problems, one world within which to work on them, and one set of standards for their solution, paradigm competition might be settled more or less routinely by some process like counting the number of problems solved by each. But, in fact, these conditions are never met completely. The proponents of *competing paradigms* are always at least slightly at cross purposes . . . the proponents will often disagree about the list of problems that any candidate for paradigm must resolve."³³

In this passage Kuhn runs together two issues which it is well to separate: one concerns the question (just addressed in the preceding section) about whether scientists have different standards of explanation or solution; the other (and the one that concerns us here) is the claim that scientists working in different paradigms want to solve different problems and that, because they do, their appraisals of the merits of theories will typically differ. So we must here deal with the case where scientists have the same standards for what counts as solving a problem but where they disagree about which problems are the most important to solve. As Kuhn puts it, "scientific controversies between the advocates of rival paradigms involve the question: which problems is it more significant to have solved? Like the issue of competing standards, that question of values can be answered only in terms of criteria that lie outside of normal science altogether."³⁴ Kuhn is surely right to insist that partisans of different global theories or paradigms often disagree about which problems it is most important to solve. But the existence of such disagreement does not establish that interparadigmatic debate about the epistemic support of rival paradigms is inevitably inconclusive or that it must be resolved by factors that lie outside the normal resources of scientific inquiry.

At first glance, Kuhn's argument seems very plausible: the differing weights assigned to the solution of specific problems by the advocates of rival paradigms may apparently lead to a situation in which the advocates of rival paradigms can each assert that their respective paradigms are the best because they solve precisely those problems they respectively believe to be the most important. No form of reasoning, insists Kuhn, could convince either side of the merits of the opposition or of the weakness of its own approach in such circumstances.

To see where Kuhn's argument goes astray in this particular instance, we need to dissect it at a more basic level. Specifically, we need to distinguish two quite distinct senses in which solving a problem may be said to be important. A problem may be important to a scientist just in the sense that he is particularly curious about it. Equally, it may be important because there is some urgent social or economic reason for solving it. Both sorts of considerations may explain why a scientist regards it as urgent to solve the problem. Such concerns are clearly relevant to explaining the motivation of scientists. But these senses of problem importance have no particular epistemic or probative significance. When we are assessing the evidential support for a theory, when we are asking how well supported or well tested that theory is by the available data, we are not asking whether the theory solves problems that are socially or personally important. Importance, in the sense chiefly relevant to this discussion, is what we might call epistemic or probative importance. One problem is of greater epistemic or probative significance than another if the former constitutes a more telling test of our theories than does the latter.

So, if Kuhn's point is to be of any significance for the epistemology of science (or, what amounts to the same thing, if we are asking how beliefworthy a theory is), then we must imagine a situation in which the advocates of different paradigms assign conflicting degrees of epistemic import to the solution of certain problems. Kuhn's thesis about such situations would be, I presume, that there is no rational machinery for deciding who is right about the assignment of epistemic weight to such problems. But that seems wrongheaded, or at least unargued, for philosophers of science have long and plausibly maintained that the primary function of scientific epistemology is precisely to ascertain the (epistemic) importance of any piece of confirming or disconfirming evidence. It is not open to a scientist simply to say that solving an arbitrarily selected problem (however great its subjective significance) is of high probative value. Indeed, it is often true that the epistemically most salient problems are ones with little or no prior practical or even heuristic significance. (Consider that Brownian motion was of decisive epistemic significance in discrediting classical thermodynamics, even though such motion had little intrinsic interest prior to Einstein's showing that such motion was anomalous for thermodynamics.) The whole point of the theory of evidence is to desubjectify the assignment of evidential significance by indicating the kinds of reasons that can legitimately be given for attaching a particular degree of epistemic importance to a confirming or refuting instance. Thus, if one maintains that the ability of a theory to solve a certain problem is much more significant epistemically than its ability to solve another, one must be able to give reasons for that epistemic preference. Put differently, one has to be able to show that the probative significance of the one problem for testing theories of a certain sort is indeed greater than that of the other. He might do so by showing that the former outcome was much more

surprising than or more general than the latter. One may thus be able to motivate a claim for the greater importance of the first problem over the second by invoking relevant epistemic and methodological criteria. But if none of these options is open to him, if he can answer the question, "Why is solving this problem more important probatively than solving that one?" only by replying, in effect, "because I am interested in solving this rather than that," then he has surrendered any claim to be shaping his beliefs rationally in light of the available evidence.

We can put the point more generally: the rational assignment of any particular degree of probative significance to a problem must rest on one's being able to show that there are viable methodological and epistemic grounds for assigning that degree of importance rather than another. Once we see this, it becomes clear that the degree of empirical support which a solved problem confers on a paradigm is not simply a matter of how keenly the proponents of that paradigm want to solve the problem.

Let me expand on this point by using an example cited extensively by both Kuhn and Doppelt: the Daltonian "revolution" in chemistry. As Doppelt summarizes the Kuhnian position, "... the pre-Daltonian chemistry of the phlogiston theory and the theory of elective affinity achieved reasonable answers to a whole set of questions effectively abandoned by Dalton's new chemistry."³⁵ Because Dalton's chemistry failed to address many of the questions answered by the older chemical paradigm, Kuhn thinks that the acceptance of Dalton's approach deprived "chemistry of some actual and much potential explanatory power."³⁶ Indeed, Kuhn is right in holding that, during most of the nineteenth century, Daltonian chemists were unable to explain many things that the older chemical traditions could make sense of. On the other hand, as Kuhn stresses, Daltonian chemistry could explain a great deal that had eluded earlier chemical theories. In short, "the two paradigms seek to explain different kinds of observational data, in response to different agendas of problems."³⁷ This "loss" of solved problems during transitions from one major theory to another is an important insight of Kuhn's. . . . But this loss of problem-solving ability through paradigm change, although real enough, does not entail, as Kuhn claims, that proponents of old and new paradigms will necessarily be unable to make congruent assessments of how well tested or well supported their respective paradigms are.

What leads Kuhn and Doppelt to think otherwise is their assumption that the centrality of a problem on one's explanatory agenda necessarily entails one's assigning a high degree of epistemic or probative weight to that problem when it comes to determining how well supported a certain theory or paradigm is. But that assumption is usually false. In general, the observations to which a reasonable scientist attaches the most probative or epistemic weight are those instances that test a theory especially "severely" (to use Popper's splendid term). The instances of greatest probative weight in the history of science (e.g., the oblate shape of the "spherical" earth,

the Arago disk experiment, the bending of light near the sun, the recession of Mercury's perihelion, the reconstitution of white light from the spectrum) have generally not been instances high on the list of problems that scientists developed their theories to solve. A test instance acquires high probative weight when, for example, it involves testing one of a theory's surprising or counterintuitive predictions, or when it represents a kind of crucial experiment between rival theories. The point is that a problem or instance does not generally acquire great probative strength in testing a theory simply because the advocates of that theory would like to be able to solve the problem. Quite the reverse, many scientists and philosophers would say. After all, it is conventional wisdom that a theory is not very acutely tested if its primary empirical support is drawn from the very sort of situations it was designed to explain. Most theories of experimental design urge—in sharp contrast with Kuhn—that theories should not be given high marks simply because they can solve the problems they were invented to solve. In arguing that the explanatory agenda a scientist sets for himself automatically dictates that scientist's reasoned judgments about well-testedness, Kuhn and Doppelt seem to have profoundly misconstrued the logic of theory appraisal.

Let us return for a moment to Kuhn's Dalton example. If I am right, Dalton might readily have conceded that pre-Daltonian chemistry solved a number of problems that his theory failed to address. Judged as theories about the qualitative properties of chemical reagents, those theories could even be acknowledged as well supported of *their type*. But Dalton's primary interests lie elsewhere, for he presumably regarded those earlier theories as failing to address what he considered to be the central problems of chemistry. But this is not an epistemic judgment; it is a pragmatic one. It amounts to saying: "These older theories are well-tested and reliable theories for explaining certain features of chemical change; but those features happen not to interest me very much." In sum, Kuhn and Doppelt have failed to offer us any grounds for thinking that a scientist's judgment about the degree of evidential support for a paradigm should or does reflect his personal views about the problems he finds most interesting. That, in turn, means that one need not share an enthusiasm for a certain paradigm's explanatory agenda in order to decide whether the theories that make up that paradigm are well tested or ill tested. It appears to me that what the Kuhn-Doppelt point really amounts to is the truism that scientists tend to invest their efforts exploring paradigms that address problems those scientists find interesting. That is a subjective and pragmatic matter which can, and should, be sharply distinguished from the question whether one paradigm or theory is better tested or better supported than its rivals. Neither Kuhn nor Doppelt has made plausible the claim that, because two scientists have different degrees of interest in solving different sorts of problems, it follows that their epistemic judgments of which theories are well tested and which are not will necessarily differ.

We are thus in a position to conclude that the existence of conflicting views among scientists about which problems are interesting apparently entails nothing about the *incompatibility* or *incommensurability* of the epistemic appraisals those scientists will make. That in turn means that these real differences of problem-solving emphasis between advocates of rival paradigms do nothing to undermine the viability of a methodology of comparative theory assessment, insofar as such a methodology is epistemically rather than pragmatically oriented. It seems likely that Kuhn and Doppelt have fallen into this confusion because of their failure to see that acknowledged differences in the motivational appeal of various problems to various scientists constitutes no rationale for asserting the existence of correlative differences in the probative weights properly assigned to those problems by those same scientists.

The appropriate conclusion to draw from the features of scientific life to which Kuhn and Doppelt properly direct our attention is that the pursuit of (and doubtless the recruitment of scientists into) rival paradigms is influenced by pragmatic as well as by epistemic considerations. That is an interesting thesis, and probably a sound one, but it does nothing to undermine the core premise of scientific epistemology: that there are principles of empirical or evidential support which are neither paradigm-specific, hopelessly vague, nor individually idiosyncratic. More important, these principles are sometimes sufficient to guide our preferences unambiguously.³⁸

■ | Notes

1. Alan Musgrave spoke for many of Kuhn's readers when he noted, apropos of the second edition of *The Structure of Scientific Revolutions*, that in "his recent writings, then, Kuhn disowns most of the challenging ideas ascribed to him by his critics . . . the new, more real Kuhn who emerges . . . [is] but a pale reflection of the old, revolutionary Kuhn" (Musgrave, 1980, p. 51).

2. It has been insufficiently noted just how partial Kuhn's break with positivism is, so far as cognitive goals and values are concerned. As I show in detail below, most of his problems about the alleged incomparability of theories arise because Kuhn accepts without argument the positivist claim that cognitive values or standards at the top of the hierarchy are fundamentally immune to rational negotiation.

3. Kuhn, 1962.

4. *Ibid.*

5. *Ibid.*, p. 159.

6. As Kuhn himself remarks, he has been attempting "to eliminate misunderstandings for which my own past rhetoric is doubtless partially responsible" (1970, pp. 259–260).

7. Kuhn, 1962, p. 108 [p. 100, above]; my italics.

8. Ibid., pp. 108–109 [p. 100, above].

9. Ibid., p. 149.

10. See [Larry Laudan, *Science and Values* (Berkeley: University of California Press, 1984), p. 43].

11. See Whewell's remarkably insightful essay of 1851, where he remarks, apropos the transition from one global theory to another: "the change . . . is effected by a transformation, or series of transformations, of the earlier hypothesis, by means of which it is brought nearer and nearer to the second [i.e., later]" (1851, p. 139).

12. Some amplification of this point is required. Kuhn evidently believes that there are some values that transcend specific paradigms. He mentions such examples as the demand for accuracy, consistency, and simplicity. The fortunes of these values are not linked to specific paradigms. Thus, if they were to change, such change would presumably be independent of shifts in paradigms. In Kuhn's view, however, these values have persisted unchanged since the seventeenth century. Or, rather, scientists have invoked these values persistently since that time; strictly speaking, on Kuhn's analysis, these values are changing constantly, since each scientist interprets them slightly differently. For a detailed discussion of Kuhn's handling of these quasi-shared values, see the final section of this chapter.

13. See Laudan, 1981.

14. For a discussion of the difference between explicit and implicit methodology, see [Laudan, *Science and Values*,] chap. 3, pp. 53 ff.

15. For an extensive bibliography on this issue, see Laudan, 1968.

16. See Laudan, 1981.

17. See [*Science and Values*, ch. 1].

18. See Laudan, [1990].

19. See *ibid.* for a lengthy treatment of some issues surrounding underdetermination of theories.

20. Kuhn, 1963, p. 152.

21. Kuhn, 1977, p. 322 [p. 103, above].

22. Ibid., p. 262.

23. Ibid., p. 329 [p. 109, above].

24. Ibid., p. 325; see also p. 324 [pp. 106, 105 above].

25. Ibid., p. 325 [p. 106, above].

26. Ibid., p. 323 [p. 105, above].

27. Ibid., p. 322 [p. 103, above].

28. Kuhn, 1970, p. 262.

29. "What I have said so far is primarily simply descriptive of what goes on in the sciences at times of theory choice" (Kuhn, 1977, p. 325 [p. 106, above]).

30. Kuhn, 1977, p. 325 [p. 106, above].

31. Doppelt, 1978. Whereas Kuhn's own discussion of these questions in *The Structure of Scientific Revolutions* rambles considerably, Doppelt offers a succinct and perspicacious formulation of what is, or at least what should have been, Kuhn's argument. Although I quarrel with Doppelt's analysis at several important points, my own thoughts about these issues owe a great deal to his writings.

32. Kuhn, 1962, p. 105 [p. 97, above].

33. Ibid., pp. 147–148.

34. Ibid., p. 110 [p. 100, above].

35. Doppelt, 1978, p. 42.

36. Kuhn, 1962, p. 107 [p. 98, above].

37. Ibid., p. 43.

38. Even on the pragmatic level, however, it is not clear that the Doppeltian version of Kuhn's relativistic picture of scientific change will stand up, for Doppelt is at pains to deny that there can be any short-term resolution between the advocates of rival axiologies. If the arguments of [*Science and Values*, ch. 2] have any cogency, it seems entirely possible that pragmatic relativism, every bit as much as its epistemic counterpart, is question begging.

■ | References

Doppelt, Gerald (1978). "Kuhn's Epistemological Relativism: An Interpretation and Defense," *Inquiry* 21: 33–86.

Gutting, Gary, ed. (1980). *Paradigms and Revolutions*. Notre Dame, Ind.: University of Notre Dame Press.

Kuhn, Thomas (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

— (1970). "Reflections on My Critics." In I. Lakatos and A. Musgrave, eds., *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press.

— (1977). *The Essential Tension*. Chicago: University of Chicago Press.

Laudan, Larry (1968). "Theories of Scientific Method from Plato to Mach," *History of Science* 7: 1–63.

— (1981). *Science and Hypothesis*. Dordrecht: Reidel.

— (1990). "Demystifying Underdetermination." In C. Wade Savage, ed., *Minnesota Studies in the Philosophy of Science*, Volume XIV, *Scientific Theories*. Minneapolis: University of Minnesota Press. [Reprinted in chapter 3, pp. 320–53, of this volume.]

Musgrave, Alan (1980). "Kuhn's Second Thoughts." In Gutting, 1980.

Whewell, William (1851). "Of the Transformation of Hypotheses in the History of Science," *Transactions of the Cambridge Philosophical Society* 9: 139–47.

HELEN E. LONGINO

Values and Objectivity

Objectivity is a characteristic ascribed variously to beliefs, individuals, theories, observations, and methods of inquiry. It is generally thought to involve the willingness to let our beliefs be determined by "the facts" or by some impartial and nonarbitrary criteria rather than by our wishes as to how things ought to be. A specification of the precise nature of such involvement is a function of what it is that is said to be objective. In this chapter I will review some common ideas about objectivity and argue that the objectivity of science is secured by the social character of inquiry. This chapter is a first step, therefore, towards socializing cognition.

Some part of the popular reverence for science has its origin in the belief that scientific inquiry, unlike other modes of inquiry, is by its very nature objective. In the modern mythology, the replacement of a mode of comprehension that simply projects human needs and values into the cosmos by a mode that views nature at a distance and dispassionately "puts nature to the question," in the words of Francis Bacon, is seen as a major accomplishment of the maturing human intellect.¹ The development of this second mode of approaching the natural world is identified, according to this view, with the development of science and the scientific method. Science is thought to provide us with a view of the world that is objective in two seemingly quite different senses of that term. In one sense objectivity is bound up with questions about the truth and referential character of scientific theories, that is, with issues of scientific realism. In this sense to attribute objectivity to science is to claim that the view provided by science is an accurate description of the facts of the natural world as they are; it is a correct view of the objects to be found in the world and of their relations with each other. In the second sense objectivity has to do with modes of inquiry. In this sense to attribute objectivity to science is

FROM Helen Longino, *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry* (Princeton, N.J.: Princeton University Press, 1990), pp. 62–82.

to claim that the view provided by science is one achieved by reliance upon nonarbitrary and nonsubjective criteria for developing, accepting, and rejecting the hypotheses and theories that make up the view. The reliance upon and use of such criteria as well as the criteria themselves are what is called scientific method. Common wisdom has it that if science is objective in the first sense it is because it is objective in the second.

At least two things can be intended by the ascription of objectivity to scientific method. Often scientists speak of the objectivity of data. By this they seem to mean that the information upon which their theories and hypotheses rest has been obtained in such a way as to justify their reliance upon it. This involves the assumption or assurance that experiments have been properly performed and that quantitative data have not been skewed by any faults in the design of survey instruments or by systematic but uncharacteristic eccentricities in the behavior of the sample studied. If a given set of data has been objectively obtained in this sense, one is thereby licensed to believe that it provides a reliable view of the world in the first of the two senses of objectivity distinguished above. . . . While objective, that is, reliable, measurement is indeed one crucial aspect of objective scientific method,² it is not the only dimension in which questions about the objectivity of methods can arise. In ascribing (or denying) objectivity to a method we can also be concerned about the extent to which it provides means of assessing hypotheses and theories in an unbiased and unprejudiced manner.

In this chapter I will explore more deeply the nature of this second mode of scientific objectivity and its connection with the logic of discourse in the natural sciences. . . . Logical positivists have relied upon formal logic and a priori epistemological requirements as keys to developing the logical analysis of science, while their historically minded wholist critics have insisted upon the primacy of scientific practice as revealed by study of the history of science. According to the former view science does indeed appear to be, by its very nature, free of subjective influence, whereas according to the latter view, subjectivity plays a major role in theory development and theory choice. Witnesses to the debate seem to be faced with a choice between two unacceptable alternatives: a logical analysis that is historically unsatisfactory and a historical analysis that is logically unsatisfactory. This kind of dilemma suggests a debate whose participants talk past one another rather than addressing common issues. Certainly part of the problem consists in attempts to develop a comprehensive account of science on the basis either of normative logical constraints or of empirical historical considerations. My analysis makes no pretense to totality or completion. It suggests, rather, a framework to be filled-in and developed both by epistemologists whose task is to develop criteria and standards of knowledge, truth, and rational belief and by historians and sociologists whose task is to make visible those historical and institutional features of the practice of science that affect its content. . . . To make way for this

interdisciplinary framework, I begin by briefly reviewing the treatment of objectivity and subjectivity in the competing analyses of the logic of science.

■ | Objectivity, Subjectivity, and Individualism

The positivist analysis of confirmation guaranteed the objectivity of science by tying the acceptance of hypotheses and theories to a public world over whose description there can be no disagreement. Positivists allow for a subjective, nonempirical element in scientific inquiry by distinguishing between a context of discovery and a context of justification.³ The context of discovery for a given hypothesis is constituted by the circumstances surrounding its initial formulation—its origin in dreams, guesses, and other aspects of the mental and emotional life of the individual scientist. Two things should be noted here. First, these nonempirical elements are understood to be features of an individual's psychology. They are treated as randomizing factors that promote novelty rather than as beliefs or attitudes that are systematically related to the culture, social structure, or socioeconomic interests of the context within which an individual scientist works. Secondly, in the context of justification these generative factors are disregarded, and the hypothesis is considered only in relation to its observable consequences, which determine its acceptability. This distinction enables positivists to acknowledge the play of subjective factors in the initial development of hypotheses and theories while guaranteeing that their acceptance remains untainted, determined not by subjective preferences but by observed reality. The subjective elements that taint its origins are purged from scientific inquiry by the methods characteristic of the context of justification: controlled experiments, rigorous deductions, et cetera. When one is urged to be objective or "scientific," it is this reliance on an established and commonly accepted reality that is being recommended. The logical positivist model of confirmation simply makes the standard view of scientific practice more systematic and logically rigorous.

As long as one takes the positivist analysis as providing a model to which any inquiry must conform in order to be objective and rational, then to the degree that actual science departs from the model it fails to be objective and rational. As noted above with respect to evidence and inference, both the historians and philosophers who have attacked the old model and those who have defended it have at times taken this position. The only disagreement with respect to objectivity, then, seems to be over the question of whether actual, historical science does or does not realize the epistemological ideal of objectivity. Defenders of the old model have argued that science ("good science") does realize the ideal. Readers of Kuhn and Feyerabend take their arguments to show that science is not

objective, that objectivity has been fetishized by traditionalists. These authors themselves have somewhat more subtle approaches. While Kuhn has emphasized the role of such subjective factors as personality, education, and group commitments in theory choice, he also denies that his is a totally subjectivist view. . . . He suggests that values such as relative simplicity and relative problem-solving ability can and do function as nonarbitrary criteria in theory acceptance. Such values can be understood as internal to inquiry, especially by those to whom scientific inquiry just is problem solving.⁴ Feyerabend, on the other hand, has rejected the relevance to science of canons of rationality or of general criteria of theory acceptance and defends a positive role for subjectivity in science.⁵

. . . How can the contextualist analysis of evidence, with its consequent denial of any logically guaranteed independence from contextual values, be accommodated within a perspective that demands or presupposes the objectivity of scientific inquiry?

As a first step in answering this question it is important to distinguish between objectivity as a characteristic of scientific method and objectivity as a characteristic of individual scientific practitioners or of their attitudes and practices. The standard accounts of scientific method tend to conflate the two, resulting in highly individualistic accounts of knowledge. Both philosophical accounts assume that method, the process by which knowledge is produced, is the application of rules to data. The positivist or traditional empiricist account of objectivity attributes objectivity to the practitioner to the extent that she or he has followed the method. Scientific method, on this view, is something that *can* be practiced by a single individual: sense organs and the capacity to reason are all that are required for conducting controlled experiments or practicing rigorous deduction. For Kuhn and for the contextualist account sketched above rationality and deference to observational data are not sufficient to guarantee the objectivity of individuals. For Kuhn this is because these intellectual activities are carried out in the context of a paradigm assented to by the scientific community. But, although Kuhn emphasizes the communitarian nature of the sciences, the theory of meaning he developed to account for the puzzling aspects of scientific change that first drew his attention reduces that community to a solipsistic monad incapable of recognizing and communicating with other monads/communities. Kuhn's account is, thus, as individualist as the empiricist one. The contextualist account makes the exercise of reason and the interpretation of data similarly dependent on a context of assumptions. Why is it not subject to the same problems?

■ | Objectivity, Criticism, and Social Knowledge

Two shifts of perspective make it possible to see how scientific method or scientific knowledge is objective even in the contextualist account. One shift is to return to the idea of science as practice. The analysis of evidential relations outlined above was achieved by thinking about science as something that is done, that involves some form of activity on the part of someone, the scientist. Because we think the goal of the scientist's practice is knowledge, it is tempting to follow tradition and seek solutions in abstract or universal rules. Refocussing on science as practice makes possible the second shift, which involves regarding scientific method as something practiced not primarily by individuals but by social groups.

The social nature of scientific practice has long been recognized. In her essay "Perception, Interpretation and the Sciences" Marjorie Grene discusses three aspects of the social character of science.⁶ One she sees as the existence of the scientific disciplines as "social enterprises," the individual members of which are dependent on one another for the conditions (ideas, instruments, et cetera) under which they practice. Another related aspect is that initiation into scientific inquiry requires education. One does not simply declare oneself a biologist but learns the traditions, questions, mathematical and observational techniques, "the sense of what to do next," from someone who has herself or himself been through a comparable initiation and then practiced. One "enters into a world" and learns how to live in that world from those who already live there. Finally, as the practitioners of the sciences all together constitute a network of communities embedded in a society, the sciences are also among a society's activities and depend for their survival on that society's valuing what they do. Much of the following can be read as an elaboration of these three points, particularly as regards the outcome, or product, of scientific practices, namely scientific knowledge. What I wish particularly to stress is that the objectivity of scientific inquiry is a consequence of this inquiry's being a social, and not an individual, enterprise.

The application of scientific method, that is, of any subset of the collection of means of supporting scientific theory on the basis of experiential data, requires by its nature the participation of two or more individuals. Even brief reflection on the actual conditions of scientific practice shows that this is so. Scientific knowledge is, after all, the product of many individuals working in (acknowledged or unacknowledged) concert. As noted earlier, scientific inquiry is complex in that it consists of different kinds of activities. It consists not just in producing theories but also in (producing) concrete interactions with, as well as models—mechanical, electrical, and mathematical—of, natural processes. These activities are carried out by different individuals, and in this era of "big science" a single complex experiment may be broken into parts, each of which will be

charged to a different individual or group of individuals. The integration and transformation of these activities ~~into a coherent understanding of a given phenomenon~~ are a matter of social negotiations.

One might argue that this is at least in principle the activity of a single individual. But, even if we were to imagine such group efforts as individual efforts, scientific knowledge is not produced by collecting the products of such imagined individuals into one whole. It is instead produced through a process of critical emendation and modification of those individual products by the rest of the scientific community. Experiments get repeated with variations by individuals other than their originators, hypotheses and theories are critically examined, restated, and reformulated before becoming an accepted part of the scientific canon. What are known as scientific breakthroughs build, whether this is acknowledged or not, on previous work and rest on a tradition of understandings, even when the effect of the breakthrough will be to undermine those understandings.⁷

The social character of scientific knowledge is made especially apparent by the organization of late twentieth-century science, in which the production of knowledge is crucially determined by the gatekeeping of peer review. Peer review determines what research gets funded and what research gets published in the journals, that is, what gets to count as knowledge. Recent concern over the breakdown of peer review and over fraudulent research simply supports the point. The most startling study of peer review suggested that scientific papers in at least one discipline were accepted on the basis of the institutional affiliation of the authors rather than the intrinsic worth of the paper.⁸ Commentary on the paper suggested that this decision procedure might be more widespread. Presumably the reviewers using the rule assume that someone would not get a job at X institution if that person were not a top-notch investigator, and so her/his experiments must be well-done and the reasoning correct. Apart from the errors in that assumption, both the reviewer and the critic of peer review treat what is a social process as an individual process. The function of peer review is not just to check that the data seem right and the conclusions well-reasoned but to bring to bear another point of view on the phenomena, whose expression might lead the original author(s) to revise the way they think about and present their observations and conclusions. To put this another way, it is to make sure that, among other things, the authors have interpreted the data in a way that is free of their subjective preferences.

The concern over the breakdown of peer review, while directed at a genuine problem, is also exaggerated partly because of an individualist conception of knowledge construction. Peer review prior to publication is not the only filter to which results are subjected. The critical treatment *after* publication is crucial to the refining of new ideas and techniques. While institutional bias may also operate in the postpublication reception of an idea, other factors, such as the attempt to repeat an experiment or

to reconcile incompatible claims, can eventually compensate for such misplaced deference. Publication in a journal does not make an idea or result a brick in the edifice of knowledge. Its absorption is a much more complex process, involving such things as subsequent citation, use and modification by others, et cetera. Experimental data and hypotheses are transformed through the conflict and integration of a variety of points of view into what is ultimately accepted as scientific knowledge.⁹

What is called scientific knowledge, then, is produced by a community (ultimately the community of all scientific practitioners) and transcends the contributions of any individual or even of any subcommunity within the larger community.¹⁰ Once propositions, theses, and hypotheses are developed, what will become scientific knowledge is produced collectively through the clashing and meshing of a variety of points of view. The relevance of these features of the sociology of science to objectivity will be apparent shortly.

The social character of hypothesis acceptance undersees the publicity of science. This publicity has both social and logical dimensions. We are accustomed to thinking of science as a public possession or property in that it is produced for the most part by public resources—either through direct funding of research or through financial support of the education of scientists. The social processes described underscore another aspect of its publicity; it is itself a public resource—a common fund of assertions presumably established to a point beyond question. It thereby constitutes a body of putative truths that can be appealed to in defense or criticism of other claims.

From a logical point of view the publicity of science includes several crucial elements. First, theoretical assertions, hypotheses, and background assumptions are all in principle public in the sense of being generally available to and comprehensible to anyone with the appropriate background, education, and interest. Second, the states of affairs to which theoretical explanations are pegged (in evidential and explanatory relationships) are public in the sense that they are intersubjectively ascertainable. . . . This does not require a commitment to a set of theory-free, eternally acceptable observation statements but merely a commitment to the possibility that two or more persons can agree about the descriptions of objects, events, and states of affairs that enter into evidential relationships. Both features are consequences of the facts (1) that we have a common language which we use to describe our experience and within which we reason and (2) that the objects of experience which we describe and about which we reason are purported to exist independently of our seeing and thinking about them.¹¹

These two aspects of the logical publicity of science make criticism of scientific hypotheses and theories possible in a way that is not possible, for instance, for descriptions of mystical experience or expressions of feeling or emotion. First, a common language for the description of experi-

ence means that we can understand each other, which means in turn that we can accept or reject hypotheses, formulate and respond to objections to them. Second, the presupposition of objects existing independently of our perception of them imposes an acceptance of constraints on what can be said or reasonably believed about them. Such acceptance implies the relevance of reports and judgments other than our own to what we say or believe. There is no way, by contrast, to acquire the authority sufficient to criticize the description of a mystical experience or the expression of a particular feeling or emotion save by having the experience or emotion in question, and these are not had in the requisite sense by more than one person. By contrast, the logical publicity of scientific understanding and subject matter makes them and hence the authority to criticize their articulation accessible to all.¹² It should be said that these constitute necessary but not sufficient conditions for the possibility of criticism, a point I shall return to later. It is the possibility of intersubjective criticism, at any rate, that permits objectivity in spite of the context dependence of evidential reasoning. Before developing this idea further let me outline some of the kinds of criticism to be found in scientific discourse.

There are a number of ways to criticize a hypothesis. For the sake of convenience we can divide these into evidential and conceptual criticism to reflect the distinction between criticism proceeding on the basis of experimental and observational concerns and that proceeding on the basis of theoretical and metatheoretical concerns.¹³ Evidential criticism is familiar enough: John Maddox, editor of *Nature*, criticizing Jacques Benveniste's experiments with highly diluted antibody solutions suggesting that immune responses could be triggered in the absence of even one molecule of the appropriate antibody;¹⁴ Richard Lewontin analyzing the statistical data alleged to favor Jensen's hypothesis of the genetic basis of I.Q.;¹⁵ Stephen Gould criticizing the experiments of David Barash purporting to demonstrate punitive responses by male mountain bluebirds to putative adultery on the part of their female mates.¹⁶ Such criticism questions the degree to which a given hypothesis is supported by the evidence adduced for it, questions the accuracy, extent, and conditions of performance of the experiments and observations serving as evidence, and questions their analysis and reporting.¹⁷

Conceptual criticism, on the other hand, often stigmatized as "metaphysical," has received less attention in a tradition of discourse dominated by empiricist ideals. At least three sorts can be distinguished. The first questions the conceptual soundness of a hypothesis—as Einstein criticized and rejected the discontinuities and uncertainties of the quantum theory;¹⁸ as Kant criticized and rejected, among other things, the Newtonian hypotheses of absolute space and time, a criticism that contributed to the development of field theory.¹⁹ A second sort of criticism questions the consistency of a hypothesis with accepted theory—as traditionalists rejected the heliocentric theory because its consequences seemed inconsis-

tent with the Aristotelian physics of motion still current in the fifteenth and sixteenth centuries;²⁰ as Millikan rejected Ehrenhaft's hypothesis of subelectrons on the basis not only of Millikan's own measurements but of his commitment to a particulate theory of electricity that implied the existence of an elementary electric charge.²¹ A third sort questions the relevance of evidence presented in support of a hypothesis: relativity theorists could deny the relevance of the Michelson-Morley interferometer experiment to the Lorentz-Fitzgerald contraction hypothesis by denying the necessity of the ether;²² Thelma Rowell and others have questioned the relevance of certain observations of animal populations to claims about dominance hierarchies within those populations by criticizing the assumptions of universal male dominance underlying claims of such relevance;²³ critics of hypotheses about the hazards of exposure to ionizing radiation direct their attention to the dose-response model with which results at high exposures are projected to conditions of low exposures.²⁴ Thus most of the debate centers not on the data but on the assumptions in light of which the data are interpreted. This last form of criticism, though related to evidential considerations, is grouped with the forms of conceptual criticism because it is concerned not with how accurately the data has been measured and reported but with the assumptions in light of which that data is taken to be evidence for a given hypothesis in the first place. Here it is not the material presented as evidence itself that is challenged but its relevance to a hypothesis.

All three of these types of criticism are central to the development of scientific knowledge and are included among the traditions of scientific discourse into which the novice is initiated. It is the third type of criticism, however, which amounts to questioning the background beliefs or assumptions in light of which states of affairs become evidence, that is crucial for the problem of objectivity. Objectivity in the sense under discussion requires a way to block the influence of subjective preference at the level of background beliefs. While the possibility of criticism does not totally eliminate subjective preference either from an individual's or from a community's practice of science, it does provide a means for checking its influence in the formation of "scientific knowledge." Thus, even though background assumptions may not be supported by the same kinds of data upon which they confer evidential relevance to some hypothesis, other kinds of support can be provided, or at least expected.²⁵ And in the course of responding to criticism or providing such support one may modify the background assumption in question. Or if the original proponent does not, someone else may do so as a way of entering into the discourse. Criticism is thereby transformative. In response to criticism, empirical support may be forthcoming (subject, of course, to the limitations developed above). At other times the support may be conceptual rather than empirical. Discussions of the nature of human judgment and cognition and whether they can be adequately modelled by computer programs, and of

the relation of subjectively experienced psychological phenomena to brain processes, for instance, are essential to theoretical development in cognitive science and neuropsychology respectively. But these discussions involve issues that are metaphysical or conceptual in nature and that, far from being resolvable by empirical means, must be resolved (explicitly or implicitly) in order to generate questions answerable by such means. The contextual analysis of evidential relations shows the limits of purely empirical considerations in scientific inquiry. Where precisely these limits fall will differ in different fields and in different research programs.

As long as background beliefs can be articulated and subjected to criticism from the scientific community, they can be defended, modified, or abandoned in response to such criticism. As long as this kind of response is possible, the incorporation of hypotheses into the canon of scientific knowledge can be independent of any individual's subjective preferences. Their incorporation is, instead, a function in part of the assessment of evidential support. And while the evidential relevance to hypotheses of observations and experiments is a function of background assumptions, the adoption of these assumptions is not arbitrary but is (or rather can be) subject to the kinds of controls just discussed. This solution incorporates as elements both the social character of the production of knowledge and the public accessibility of the material with which this knowledge is constructed.

Sociologically and historically, the molding of what counts as scientific knowledge is an activity requiring many participants. Even if one individual's work is regarded as absolutely authoritative over some period—as for instance, Aristotle's and later Newton's were—it is eventually challenged, questioned, and made to take the role of contributor rather than sole author—as Aristotle's and Newton's have been. From a logical point of view, if scientific knowledge were to be understood as the simple sum of finished products of individual activity, then not only would there be no way to block or mitigate the influence of subjective preference but scientific knowledge itself would be a potpourri of merrily inconsistent theories. Only if the products of inquiry are understood to be formed by the kind of critical discussion that is possible among a plurality of individuals about a commonly accessible phenomenon, can we see how they count as knowledge rather than opinion.

Objectivity, then, is a characteristic of a community's practice of science rather than of an individual's, and the practice of science is understood in a much broader sense than most discussions of the logic of scientific method suggest. These discussions see what is central to scientific method as being the complex of activities that constitute hypothesis testing through comparison with experiential data—in principle, if not always in reality, an activity of individuals. What I have argued here is that scientific method involves as an equally central aspect the subjection of hypotheses and the background assumptions in light of which they seem

to be supported by data to varieties of conceptual criticism, which is a social rather than an individual activity.²⁶

The respect in which science is objective, on this view, is one that it shares with other modes of inquiry, disciplines such as literary or art criticism and philosophy.²⁷ The feature that has often been appealed to as the source of the objectivity of science, that its hypotheses and theories are accepted or rejected on the basis of observational, experimental data, is a feature that makes scientific inquiry empirical. In the positivist account, for instance, it was the syntactically and deductively secured relation of hypotheses to a stable set of observational data that guaranteed the objectivity of scientific inquiry. But, as I've argued, most evidential relations in the sciences cannot be given this syntactic interpretation. In the contextual analysis of evidential relations, however, that a method is empirical in the above sense does not mean that it is also objective. A method that involved the appeal to observational or experimental data but included no controls on the kinds of background assumptions in light of which their relevance to hypotheses might be determined, or that permitted a weekly change of assumptions so that a hypothesis accepted in one week on the basis of some bit of evidence *e* would be rejected the next on the same basis, would hardly qualify as objective. Because the relation between hypotheses and evidence is mediated by background assumptions that themselves may not be subject to empirical confirmation or disconfirmation, and that may be infused with metaphysical or normative considerations, it would be a mistake to identify the objectivity of scientific methods with their empirical features alone. The process that can expose such assumptions is what makes possible, even if it cannot guarantee, independence from subjective bias, and hence objectivity. Thus, while rejecting the idea that observational data alone provide external standards of comparison and evaluation of theories, this account does not reject external standards altogether. The formal requirement of demonstrable evidential relevance constitutes a standard of rationality and acceptability independent of and external to any particular research program or scientific theory. The satisfaction of this standard by any program or theory, secured, as has been argued, by intersubjective criticism, is what constitutes its objectivity.

Scientific knowledge is, therefore, social knowledge. It is produced by processes that are intrinsically social, and once a theory, hypothesis, or set of data has been accepted by a community, it becomes a public resource. It is available to use in support of other theories and hypotheses and as a basis of action. Scientific knowledge is social both in the ways it is created and in the uses it serves.

■ | Objectivity by Degrees

I have argued both that criticism from alternative points of view is required for objectivity and that the subjection of hypotheses and evidential reasoning to critical scrutiny is what limits the intrusion of individual subjective preference into scientific knowledge. Are these not two opposing forms of social interaction, one dialogic and the other monologic? Why does critical scrutiny not simply suppress those alternative points of view required to prevent premature allegiance to one perspective? How does this account of objectivity not collapse upon itself? The answer involves seeing dialogic and monologic as poles of a continuum. The maintenance of dialogue is itself a social process and can be more or less fully realized. Objectivity, therefore, turns out to be a matter of degree. A method of inquiry is objective to the degree that it permits *transformative* criticism. Its objectivity consists not just in the inclusion of intersubjective criticism but in the degree to which both its procedures and its results are responsive to the kinds of criticism described. I've argued that method must, therefore, be understood as a collection of social, rather than individual, processes, so the issue is the extent to which a scientific community maintains critical dialogue. Scientific communities will be objective to the degree that they satisfy four criteria necessary for achieving the transformative dimension of critical discourse: (1) there must be recognized avenues for the criticism of evidence, of methods, and of assumptions and reasoning; (2) there must exist shared standards that critics can invoke; (3) the community as a whole must be responsive to such criticism; (4) intellectual authority must be shared equally among qualified practitioners. Each of these criteria requires at least a brief gloss.

RECOGNIZED AVENUES FOR CRITICISM

The avenues for the presentation of criticism include such standard and public forums as journals, conferences, and so forth. Peer review is often pointed to as the standard avenue for such criticism, and indeed it is effective in preventing highly idiosyncratic values from shaping knowledge. At the same time its confidentiality and privacy make it the vehicle for the entrenchment of established views. This criterion also means that critical activities should receive equal or nearly equal weight to "original research" in career advancement. Effective criticism that advances understanding should be as valuable as original research that opens up new domains for understanding; pedestrian, routine criticism should be valued comparably to pedestrian and routine "original research."

SHARED STANDARDS

In order for criticism to be relevant to a position it must appeal to something accepted by those who hold the position criticized. Similarly, alternative theories must be perceived to have some bearing on the concerns of a scientific community in order to obtain a hearing. This cannot occur at the whim of individuals but must be a function of public standards or criteria to which members of the scientific community are or feel themselves bound. These standards can include both substantive principles and epistemic, as well as social, values. Different subcommunities will subscribe to different but overlapping subsets of the standards associated with a given community. Among values the standards can include such elements as empirical adequacy, truth, generation of specifiable interactions with the natural or experienced world, the expansion of existing knowledge frameworks, consistency with accepted theories in other domains, comprehensiveness, reliability as a guide to action, relevance to or satisfaction of particular social needs. Only the first of these constitutes a necessary condition that any research program must meet or aspire to meet, and even this requirement may be temporarily waived and is subject to interpretation.

The list shares some elements with the list Thomas Kuhn presents in his essay "Objectivity, Value Judgment, and Theory Choice,"²⁸ and like the items in his list they can be weighted differently in different scientific communities and they must be more precisely formulated to be applicable. For example, the requirement that theories have some capability to generate specifiable interactions with the natural or experienced world will be applied differently as the sorts of interactions desired in a community differ. The particular weighting and interpretation assigned these standards will vary in different social and historical contexts as a function of cognitive and social needs. Furthermore, they are not necessarily consistent. . . . The goals of truth or accurate representation and expansion of existing knowledge frameworks exist in some tension with each other.

Standards do not provide a deterministic theory of theory choice. Nevertheless, it is the existence of standards that makes the individual members of a scientific community responsible to something besides themselves. It is the open-ended and nonconsistent nature of these standards that allows for pluralism in the sciences and for the continued presence, however subdued, of minority voices. Implicit or explicit appeals to such standards as I've listed underwrite many of the critical arguments named above.

COMMUNITY RESPONSE

This criterion requires that the beliefs of the scientific community as a whole and over time change in response to the critical discussion taking

place within it. This responsiveness is measured by such public phenomena as the content of textbooks, the distribution of grants and awards, the flexibility of dominant world views. Satisfaction of this criterion does not require that individuals whose data and assumptions are criticized recant. Indeed, understanding is enhanced if they can defend their work against criticism.²⁹ What is required is that community members pay attention to the critical discussion taking place and that the assumptions that govern their group activities remain logically sensitive to it.

EQUALITY OF INTELLECTUAL AUTHORITY

This Habermasian criterion is intended to disqualify a community in which a set of assumptions dominates by virtue of the political power of its adherents.³⁰ An obvious example is the dominance of Lamarckism in the Soviet Union in the 1930s. While there were some good reasons to try experiments under the aegis of a Lamarckian viewpoint, the suppression of alternative points of view was a matter of politics rather than of logic or critical discussion. The bureaucratization of United States science in the twentieth century tends similarly to privilege certain points of view.³¹ The exclusion, whether overt or more subtle, of women and members of certain racial minorities from scientific education and the scientific professions has also constituted a violation of this criterion. While assumptions about race and about sex are not imposed on scientists in the United States in the way assumptions about inheritability of acquired traits were in the Soviet Union, . . . assumptions about sex structure a number of research programs in biology and behavioral sciences. Other scholars have documented the role of racial assumptions in the sciences.³² The long-standing devaluation of women's voices and those of members of racial minorities means that such assumptions have been protected from critical scrutiny.

The above are criteria for assessing the objectivity of communities. The objectivity of individuals in this scheme consists in their participation in the collective give-and-take of critical discussion and not in some special relation (of detachment, hardheadedness) they may bear to their observations. Thus understood, objectivity is dependent upon the depth and scope of the transformative interrogation that occurs in any given scientific community. This communitywide process ensures (or can ensure) that the hypotheses ultimately accepted as supported by some set of data do not reflect a single individual's idiosyncratic assumptions about the natural world. To say that a theory or hypothesis was accepted on the basis of objective methods does not entitle us to say it is true but rather that it reflects the critically achieved consensus of the scientific community. In the absence of some form of privileged access to transempirical (unobservable) phenomena it's not clear we should hope for anything better.

The weight given to criticism in the formation of knowledge represents a social consensus regarding the appropriate balance between ac-

curate representation and knowledge extension. Several conditions can limit the extent of criticism and hence diminish a scientific community's objectivity without resulting in a completely or intentionally closed society (for example, such as characterized Soviet science under Stalin or some areas of Nazi science).

First of all, if scientific inquiry is to have any effect on a society's ability to take advantage of natural processes for the improvement of the quality of its life, criticism of assumptions cannot go on indefinitely. From a logical point of view, of course, criticism of background assumptions, as of any general claim, can go on *ad infinitum*. The philosophical discussion of inductive reasoning is an example of such unending (though not useless) debate. The utility of scientific knowledge depends on the possibility of finding frameworks of inquiry that remain stable enough to permit systematic interactions with the natural world. When critical discussion becomes repetitive and fixed at a metalevel, or when criticism of one set of assumptions ceases to have or does not eventually develop a connection to an empirical research program, it loses its relevance to the construction of empirical knowledge. It is the intrinsic incapacity of so-called "creation science" to develop a fruitful research program based on its alleged alternative to evolutionary theory that is responsible for the lack of attention given to it by the contemporary United States scientific community. The appeal by its advocates to pluralistic philosophies of science seems misguided, if not disingenuous.

Secondly, these critical activities, however crucial to knowledge building, are de-emphasized in a context that rewards novelty and originality, whether of hypotheses or of experimental design. The commoditization of scientific knowledge—a result of the interaction of the requirements of career advancement and of the commercial value of data—diminishes the attention paid to the criticism of the acquisition, sorting, and assembling of data. It is a commonplace that in contemporary science papers reporting negative results do not get published.

In the third place, some assumptions are not perceived as such by any members of the community. When, for instance, background assumptions are shared by all members of a community, they acquire an invisibility that renders them unavailable for criticism. They do not become visible until individuals who do not share the community's assumptions can provide alternative explanations of the phenomena without those assumptions, as, for example, Einstein could provide an alternative explanation of the Michelson-Morley interferometer experiment. Until such alternatives are available, community assumptions are transparent to their adherents. In addition, the substantive principles determining standards of rationality within a research program or tradition are for the most part immune to criticism by means of those standards.

From all this it follows again that the greater the number of different points of view included in a given community, the more likely it is that

its scientific practice will be objective, that is, that it will result in descriptions and explanations of natural processes that are more reliable in the sense of less characterized by idiosyncratic subjective preferences of community members than would otherwise be the case. The smaller the number, the less likely this will be.³³ Because points of view cannot simply be allowed expression but must have an impact on what is ultimately thought to be the case, such diversity is a necessary but not a sufficient condition for objectivity. Finally, these conditions reinforce the point that objectivity is a matter of degree. While the conditions for objectivity are at best imperfectly realized, they are the basis of an ideal by reference to which particular scientific communities can be evaluated. Ascertaining in greater detail the practices and institutional arrangements that facilitate or undermine objectivity in any particular era or current field, and thus the degree to which the ideal of objectivity is realized, requires both historical and sociological investigation. . . .

■ | Conclusion

On the positivist analysis of scientific method it is hard to understand how theories purporting to describe a nonobservable underlying reality, or containing descriptive terms whose meaning is independent of that of so-called observational terms, can be supported. On the antiempiricist wholist account it is just as difficult to understand how the theories that are developed have a bearing on intersubjective reality. Each of these approaches is also unable to account for certain facts about the actual practice of science. The absolute and unambiguous nature of evidential relations presented in the positivist view cannot accommodate the facts of scientific change. The incommensurability of theories in the wholist view cannot do justice to the lively and productive debate that can occur among scientists committed to different theories. Each of these modes of analysis emphasizes one aspect of scientific method at the expense of another, and each produces an individualist logic of scientific method that fails adequately to reflect the social nature of scientific discourse. Furthermore, the emphasis on theories distorts scientific growth and practice. Scientists rarely engage in the construction or evaluation of comprehensive theories. Their constructive, theoretical activity tends to consist much more in the development of individual or interrelated hypotheses (as laws, generalizations, or explanations) from the complex integration of observation and experiment with background assumptions. Success in expanding the scope of an explanatory idea via such complex integration plays as important a role in its acceptance as the survival of falsifying tests. Accounts of validation in the sciences must take account both of the role of background assumptions in evidential reasoning and of the roles of (sometimes) con-

flicting goals of inquiry with respect to which hypotheses and theories are assessed. The logic that reflects the structure of this activity will have to abandon some of the simplicity of the positivist account, but what it loses in elegance it will surely regain in application.

The analysis conducted in this chapter means that values can enter into theory-constructive reasoning in two major ways—through an individual's values or through community values. The fact that a bit of science can be analyzed as crucially dependent on contextual values or on value-laden background assumptions does not necessarily mean that someone is attempting to impose his/her wishes on the natural world without regard to what it might really be like. More customarily such analysis should be taken as showing the way in which such contextual features have facilitated the use of given data or observations as evidence for some hypothesis by an individual or by a community. Because community values and assumptions determine whether a given bit of reasoning will pass or survive criticism and thus be acceptable, individual values as such will only rarely be at issue in these analyses. When an individual researcher's values enable her or him to make inferences at variance with those of the scientific community, this is less evidence of strongly eccentric individualism than of allegiance to some other social (political or religious) community.³⁴

The contextualist view produces a framework within which it is possible to respect the complexity of science, to do justice to the historical facts and to the current practice of science, and to avoid paradox. In addition, it is possible to articulate a standard of comparison independent of and external to any particular theory or research project. In making inter-theoretic comparison possible it offers the basis (an expanded basis) upon which to develop criteria of evaluation. Finally, the social account of objectivity and scientific knowledge to which the contextualist account of evidence leads seems more true to the fact that scientific inquiry is not always as free from subjective preference as we would wish it to be. And even though the resulting picture of objectivity differs from what we are used to, our intuition that scientific inquiry at its best is objective is kept intact by appealing to the spirit of criticism that is its traditional hallmark.³⁵

■ | Notes

1. This mythology originates with the founders of modern science—compare Isaac Newton's "Rules of Reasoning in Philosophy" in Newton (1953), pp. 3–5—and has come to be the standard view.

2. It has become a subject of increased concern lately in light of several alleged incidents of data faking. Compare Broad (1981).

3. Hempel (1966), pp. 3–18, and Popper (1962), pp. 42–59.

4. Laudan (1977) does articulate criteria for what counts as progress. These are not necessarily criteria or standards for truth.

5. Feyerabend (1975).

6. Grene (1985).

7. James Watson's account of the discovery of the molecular structure of DNA, read in conjunction with the story of Rosalind Franklin's contributions to that discovery in Sayre (1975), provides a vivid example of this interdependence. See Watson (1968). Participant accounts of recent developments in one or another science usually offer good illustrations of this point. Weinberg (1977) and Feinberg (1978) account for the mid-1970s states of cosmology and microphysics, respectively. Each presents what can be called the current canon in its field, making clear the dependence of its production upon the activity and interaction of many individual researchers.

8. See Peters and Ceci (1982, 1985) and the associated commentary. For additional discussion of peer review see Glazer (1988); Goleman (1987); Cole and Cole (1977); Cole, Cole, and Simons (1981).

9. In what I take to be a similar vein, Bruno Latour (1987) claims that in science a statement made by an individual becomes a fact only as a consequence of what others do with the statement. Latour, however, emphasizes the agonistic as opposed to the cooperative dimension of social relations in the sciences.

10. The precise extension of "scientific community" is here left unspecified. If it includes those interested in and affected by scientific inquiry, then it is much broader than the class of those professionally engaged in scientific research. For a discussion of these issues and some consequences of our current restricted understanding of the scientific community see Addelson (1983).

11. One might say that the language game of science presupposes the independent existence of objects of experience. Contemporary arguments about scientific realism can be understood as arguments about (1) the nature of this presupposition and (2) what categories of objects it covers.

12. To avoid possible confusion about the point being made here, I wish to emphasize that I am contrasting the descriptive statements of science with expressions of emotion. *Descriptions* of emotion and other subjective states may be as objective as other kinds of description, if the conditions for objectivity can be satisfied. Objectivity as it is being discussed here involves the absence (or control) of subjective *preference* and is not necessarily divorced from our beliefs about our subjective states. Locke (1968) discusses the different ways in which privacy is properly and improperly attributed to subjective states (pp. 5–12).

13. The distinction between the different kinds of concerns relevant to the development and evaluation of theories is discussed for different purposes and with significant differences in detail by Buchdahl in a discussion of criteria choice, by Laudan in a discussion of the problems that give rise to the development of theory, and by Schaffner in a discussion of categories for comparative theory evaluation. A more complete categorization of concerns and types of criticism than that offered here requires a more thorough study of past and present scientific practice. See Gerd Buchdahl (1970); Larry Laudan (1977), and Kenneth Schaffner (1974).

14. Maddox, Randi, and Stewart (1988) and Benveniste's reply in Benveniste (1988). The chapter "Laboratories" in Latour (1987) can be read as providing a series of examples of evidential criticism (pp. 63–100).

15. Lewontin (1970, 1974).

16. Gould (1980).

17. The latter two kinds of questions are concerned with the objectivity of data, a notion mentioned above.

18. Bernstein (1973), pp. 137–177.

19. Williams (1966), pp. 32–63. A somewhat different account is presented by Hesse (1965), pp. 170–180.

20. Kuhn (1957), pp. 100–133, 185–192.

21. Holton (1978).

22. Jaffe (1960), pp. 95–103.

23. Rowell (1974).

24. See Longino (1987).

25. Conceptual criticism of this sort is a far cry from the criticism envisaged by Popper. For him metaphysical issues must be decided empirically, if at all. (And if they cannot be so tested, they lack significance.)

26. This is really a distinction between the number of points of view (minds) required. Many individuals (sharing assumptions and points of view) may be involved in testing a hypothesis (and commonly are in contemporary experiments). And though this is much rarer, one individual may be able to criticize her or his own evidential reasoning and background assumptions from other points of view.

27. This is not to deny the importance of distinguishing between different modes of understanding—for instance, between scientific, philosophical, and literary theories—but simply to deny that objectivity can serve as any kind of demarcation criterion.

28. Kuhn (1977) [pp. 102–18, above].

29. Beatty (1985) makes a similar point.

30. Invocation of this criterion confirms the kinship of this account of objectivity with the account of truth that Jürgen Habermas has developed as part of his theory of communicative competence. . . .

31. See Levins and Lewontin (1985), pp. 197–252, for further discussion of this point.

32. See Gould (1981); Lewontin, Rose, and Kamin (1984); Richardson (1984).

33. This insistence on the variety of points of view required for objectivity is developed on a somewhat different basis for the social sciences by Sandra Harding (1978).

34. This should not be taken to mean that social inequality and marginalization are necessary for objectivity but rather that differences in perspective are. A scientific community existing in a (utopian at this point) society characterized by

thoroughgoing inclusivity and equality might indeed encourage the persistence of divergent points of view to ensure against blindness to its own assumptions.

35. Note added in proof. Three books read since completing the manuscript also draw attention in varying degrees to the social character of cognitive processes in science: Peter Galison, *How Experiments End* (Chicago, IL: University of Chicago Press, 1987); David Hull, *Science as a Process* (Chicago, IL: University of Chicago Press, 1988); and Sharon Traweek, *Beamtimes and Lifetimes: The World of High Energy Physicists* (Cambridge, MA: Harvard University Press, 1988).

References

- Addelson, Kathryn Pyne. 1983. "The Man of Professional Wisdom." In *Discovering Reality: Feminist Perspectives on Epistemology, Metaphysics, Methodology, and Philosophy of Science*, ed. Sandra Harding and Merrill Hintikka, pp. 165–186. Dordrecht: D. Reidel.
- Beatty, John. 1985. "Pluralism and Panselctionism." In *PSA 1984*, ed. Peter Asquith and Philip Kitcher, pp. 25–83. East Lansing, MI: Philosophy of Science Association.
- Benveniste, Jacques. 1988. "Reply to Maddox, Randi and Stewart." *Nature* 334: 291.
- Bernstein, Jeremy. 1973. *Einstein*. Bungay: William Collins and Son, Ltd.
- Broad, William. 1981. "Fraud and the Structure of Science." *Science* 212: 137–141.
- Buchdahl, Gerd. 1970. "History of Science and Criteria of Choice." In *Minnesota Studies in the Philosophy of Science*, ed. Roger Steuwer, 5: 205–230. Minneapolis: University of Minnesota Press.
- Cole, Stephen, Jonathan R. Cole, and Gary Simons. 1981. "Chance and Consensus in Peer Review." *Science* 214: 881–886.
- Cole, Stephen, Leonard Rubin, and Jonathan R. Cole. 1977. "Peer Review and the Support of Science." *Scientific American* 237, no. 4: 34–41.
- Feinberg, Gerald. 1978. *What Is the World Made Of?* New York: Anchor Press.
- Feyerabend, Paul K. 1975. *Against Method*. London: Verso.
- Glazer, Sarah. 1988. "Combating Science Fraud." *Editorial Research Reports* 2: 390–399.
- Goleman, Daniel. 1987. "Failing to Recognize Bias in Science." *Technology Review* 90 (November–December): 26–27.
- Gould, Stephen J. 1980. "Sociobiology and the Theory of Natural Selection." In *Sociobiology: Beyond Nature/Nurture?* ed. George Barlow and James Silverberg, pp. 257–269. Boulder, CO: Westview Press.
- . 1981. *The Mismeasure of Man*. New York: W.W. Norton and Co.
- Grene, Marjorie. 1985. "Perception, Interpretation and the Sciences." In *Evolution at a Crossroads*, ed. David Depew and Bruce Weber, pp. 1–20. Cambridge, MA: MIT Press.

- Harding, Sandra. 1978. "Four Contributions Values Can Make to the Objectivity of the Social Sciences." In *Proceedings of the 1978 Biennial Meeting of the Philosophy of Science Association*, ed. Peter Asquith and Ian Hacking, pp. 199–209. East Lansing, MI: Philosophy of Science Association.
- Hempel, Carl Gustav. 1966. *Philosophy of Natural Science*. Englewood Cliffs, NJ: Prentice Hall.
- Hesse, Mary. 1965. *Forces and Fields*. Totowa, NJ: Littlefield Adams.
- Holton, Gerald. 1978. "Subelectrons, Presuppositions, and the Millikan-Ehrenhaft Dispute." In *The Scientific Imagination*, pp. 25–83. Cambridge: Cambridge University Press.
- Jaffe, Bernard. 1960. *Michelson and the Speed of Light*. Garden City, NY: Doubleday and Co.
- Kuhn, Thomas. 1957. *The Copernican Revolution*. Cambridge, MA: Harvard University Press.
- . 1977. *The Essential Tension*. Chicago: University of Chicago Press.
- Latour, Bruno. 1987. *Science in Action*. Cambridge MA: Harvard University Press.
- Laudan, Larry. 1977. *Progress and Its Problems*. Berkeley, CA: University of California Press.
- Levins, Richard, and Richard Lewontin. 1985. *The Dialectical Biologist*. Cambridge, MA: Harvard University Press.
- Lewontin, Richard. 1970. "Race and Intelligence." *Bulletin of the Atomic Scientists* 26 (March): 2–8.
- . 1974. "The Analysis of Variance and the Analysis of Causes." *American Journal of Human Genetics* 26: 400–411.
- Lewontin, Richard, Steven Rose, and Leon Kamin. 1984. *Not in Our Genes: Biology, Ideology and Human Nature*. New York: Pantheon Books.
- Locke, Don. 1968. *Myself and Others*. London: Oxford University Press.
- Longino, Helen. 1987. "What's Really Wrong with Quantitative Risk Assessment?" In *PSA 1986*, ed. Arthur Fine and Peter Machamer, pp. 376–383. East Lansing, MI: Philosophy of Science Association.
- Maddox, John, James Randi, and Walter W. Stewart. 1988. "High Dilution Experiments a Delusion." *Nature* 334: 287–290.
- Newton, Isaac. 1953. "Rules of Reasoning in Philosophy." In *Newton's Philosophy of Nature*, ed. H. S. Thayer, pp. 3–5. New York: Hafner.
- Peters, Donald, and Stephen Ceci. 1982. "Peer Review Practices of Psychological Journals: The Fate of Published Articles Submitted Again." *Behavioral and Brain Sciences* 5: 187–195.
- . 1985. "Beauty Is in the Eye of the Beholder." *Behavioral and Brain Sciences* 8, no. 4: 747–749.
- Popper, Karl. 1959. *The Logic of Scientific Discovery*. London: Hutchinson.
- . 1962. *Conjectures and Refutations*. New York: Basic Books.

Richardson, Robert C. 1984. "Biology and Ideology: The Interpenetration of Science and Values." *Philosophy of Science* 51: 396-420.

Rowell, Thelma. 1974. "The Concept of Dominance." *Behavioral Biology* 11: 131-154.

Sayre, Anne. 1975. *Rosalind Franklin and DNA*. New York: W. W. Norton.

Schaffner, Kenneth. 1974. "Einstein versus Lorentz: Research Programmes and the Logic of Theory Evaluation." *British Journal for the Philosophy of Science* 25: 43-78.

Watson, James. 1968. *The Double Helix*. New York: Atheneum.

Weinberg, Steven. 1977. *The First Three Minutes*. New York: Basic Books.

Williams, L. Pearce. 1966. *The Origins of Field Theory*. New York: Random House.

Gender and the Biological Sciences

Feminist critiques of science provide fertile ground for any investigation of the ways in which social influences may shape the content of science. Many authors working in this field are from the natural and social sciences; others are philosophers. For philosophers of science, recent work on sexist and androcentric bias in science raises hard questions about the extent to which reigning accounts of scientific rationality can deal successfully with mounting evidence that gender ideology has had deep and extensive effects on the development of many scientific disciplines.

Feminist critiques of biology have been especially important in the political struggle for gender equality because biologically determinist arguments are so often cited to 'explain' women's oppression. They explain why it is 'natural' for women to function in a socially subordinate role, why men are smarter and more aggressive than women, why women are destined to be homebodies, and why men rape. Genes, hormones, and evolutionary processes are cited as determinants of this natural order and ultimately as evidence that interventions to bring about a more egalitarian and just society are either useless or counterproductive.

The critiques of biology are also *epistemically* important because of the position that biology occupies in the usual hierarchy of the sciences—somewhere between physics, on the one hand, and the social sciences, on the other. Very often feminist critiques of the social sciences are dismissed out of hand by philosophers of science on the grounds that the social sciences aren't science anyway; and so the feminist critiques, however devastating, are said to tell us nothing about the nature of real science. It is, however, not quite so easy to dismiss biology as pseudo-science; and so the critiques in this area assume added significance. If we are to infer in light of the feminist critiques anything about the nature of science (its

rationality, its objectivity, its degree of insulation from social influences, its character as an individual or collective enterprise), then the biological sciences are perhaps the best place to start. Hence this essay. It has four parts. In section I, several case studies of gender ideology in the biological sciences are reviewed. This review provides a common stock of examples for discussion purposes and the opportunity to indicate very briefly how standard theses in philosophy of science can provide partial illumination of them. In section II, the possible epistemic significance of these case studies (and others like them) is addressed in light of alternative conceptions of science available in the feminist literature. The third part of the essay develops an account of the relation between contexts of discovery and justification that makes room for the sorts of social and cultural influences on science exemplified by gender bias while still allowing room for fairly robust notions of objectivity and rationality. Finally, in section IV, an attempt is made to locate this account vis-à-vis others represented among feminist critiques of science.

I | Some Case Studies

Consider first a 1988 article entitled 'The Importance of Feminist Critiques for Contemporary Cell Biology,' authored by the Biology and Gender Study Group at Swarthmore College.¹ The article discusses the ways in which contemporary research is still shackled by outmoded models of the relationship between egg and sperm in reproduction. In particular, commitment to the Sleeping Beauty/Prince Charming model of egg and sperm may have blinded researchers and theoreticians to some of the facts about human reproduction. Just as women are seen to be passive and men active, so traditionally have egg and sperm been assigned the traditional feminine and masculine roles. The egg waits passively while the sperm heroically battles upstream, struggles against the hostile uterus, courts the egg, and (if victorious) penetrates by burrowing through, thereby excluding all rival suitors. The egg's only role in this saga is to select which rival will be successful.

The notion that the male semen awakens the slumbering egg appeared as early as 1795 and has been influential ever since. In the last fifteen years, however, some rival accounts have challenged the old narrative by making the egg an energetic partner in fertilization. For example, using electron microscopy it can be shown that the sperm doesn't just burrow through the egg, as previously thought. Instead, the egg directs the growth of small, finger-like projections of the cell surface to *clasp* the sperm and slowly draw it in. This mound of microvilli extending to the sperm was discovered in 1895 when the first photographs of sea urchin fertilization were published; but it has largely been ignored until recently.

What matters for our purposes here is not whether the newer theory is entirely correct (it is still controversial), but that its very existence as a rival to the more established views throws into sharp relief the questionable assumptions of the older model. It shows us how pre-existing theoretical assumptions inform which questions we ask, which hypotheses we investigate, and which data we decide to ignore as evidentially insignificant. These considerations are sometimes relegated to the context of discovery and are said to be epistemically irrelevant to the actual content of science. This is a topic to which we shall return later. In the meantime, let us investigate some cases in which the controverted question is not whether some data are evidentially significant at all, but which interpretation should be placed upon the same data as the result of differing theoretical commitments.

Many feminist criticisms of primatology and sociobiology focus on the fact that male struggle, male competition, and male inventiveness are portrayed as the bases for human evolution. In familiar passages from *The Descent of Man* quoted by Ruth Hubbard and other critics, Darwin attributes evolutionary development in human beings almost exclusively to male activity.

[Men have had] to defend their females, as well as their young, from enemies of all kinds, and to hunt for their joint subsistence. But to avoid enemies or to attack them with success, to capture wild animals, and to fashion weapons, requires the aid of the higher mental faculties, namely observation, reason, invention, or imagination. These various faculties will thus have been continually put to the test and selected during manhood.

'Thus,' the discussion ends, 'man has ultimately become superior to woman' and it is a good thing that men pass on their characteristics to their daughters as well as to their sons, 'otherwise it is probable that man would have become as superior in mental endowment to woman, as the peacock is in ornamental plumage to the peahen.'²

The influence of Darwin's androcentric bias has not been limited to evolutionary biology, since that theory functions as an auxiliary hypothesis in many other disciplines. Consider, for example, anthropology. If one holds the view that man-the-hunter is chiefly responsible for human evolutionary development, one interprets fossil evidence in light of the changing behavior of males. Helen Longino and Ruth Doell, for example, in a very important 1983 article called 'Body, Bias, and Behavior: A Comparative Analysis of Reasoning in Two Areas of Biological Science,'³ trace the ways in which the androcentric account attributes the development of tool use to male hunting behavior. Some recent work, however, has suggested that up to 80% of the subsistence diet of what used to be called hunter-gatherer societies came from female gatherers. If that is the background theory informing one's interpretation of the evidence, then quite a differ-

ent account of that same evidence emerges. This is how Longino and Doell summarize the point:

By contrast [with the androcentric account], the gynecentric story explains the development of tool use as a function of female behavior, viewing it as a response to the greater nutritional stress experienced by females during pregnancy, and later in the course of feeding their young through lactation and with foods gathered from the surrounding savanna. Whereas man-the-hunter theorists focus on stone tools, woman-the-gatherer theorists see tool use developing much earlier and with organic materials such as sticks and reeds. They portray females as innovators who contributed more than males to the development of such allegedly human characteristics as greater intelligence and flexibility. Women are said to have invented the use of tools to defend against predators while gathering and to have fashioned objects to serve in digging, carrying, and food preparation.

Again, what matters here is not that the gynecentric hypothesis be true but rather that it makes obvious the extent to which the standard interpretation of the anthropological evidence has been colored by androcentric bias.

The cases examined so far are instances in which attention to the theory-ladenness of observation or the underdetermination of theory by data shed some light on the way in which pre-existing theoretical commitments regarding sex and gender may influence decisions about which questions get asked, which data must be accounted for and which can safely be ignored, as well as which interpretation among those that are empirically adequate is actually adopted. There are other cases in which attention to the Duhem-Quine thesis is helpful. Even if the body of relevant data has already been strictly delimited with preferred interpretations settled upon, and even if the test hypothesis has been selected, it is still to some extent an open question how one ought to respond to apparently falsifying information. Although one may simply reject the test hypothesis, it is also possible to pin the blame for a failed prediction on one of the background assumptions that was used to generate the failed prediction. The arrow of *modus tollens*, in other words, may be redirected away from the test hypothesis and toward one or more of the auxiliaries. This, of course, can be a perfectly respectable and useful response to failed prediction; but it does raise interesting questions about what factors (social as well as more narrowly 'cognitive') motivate our decisions to protect some hypotheses from falsification. It also draws attention to the important role played in theory assessment by our background assumptions, a role that is particularly crucial in the present discussion since so few of our background assumptions about sex and gender have been subjected to systematic scrutiny.

Certain hypotheses seem to survive one falsification after another, with

the blame for failure and the subsequent adjustment always being located elsewhere in the system of beliefs. I have in mind here recent developments in neuroanatomy which are directed to explaining intelligence differences between women and men, particularly as these relate to alleged male superiority with respect to mathematical and spatial ability. Anne Fausto-Sterling, in her book *Myths of Gender*,⁴ has surveyed some recent theories; the following examples are taken from her discussion.

It has been suggested that spatial ability is X-linked and therefore exhibited more frequently in males than in females; that high levels of prenatal androgen increase intelligence; that lower levels of estrogen lead to superior male ability at 'restructuring' tasks. Some have held that female brains are more lateralized than male brains and that greater lateralization interferes with spatial functions. Others have argued that female brains are less lateralized than male brains and that less lateralization interferes with spatial ability. Some have attempted to save the hypothesis of X-linked spatial ability from refuting evidence by suggesting that the sex-linked spatial gene can be expressed only in the presence of testosterone. Others have argued that males are smarter because they have more uric acid than females.

None of these hypotheses is well-supported by the evidence and most seem to be clearly refuted. What is interesting for our purposes is that for many researchers the one element of the theoretical network they are unwilling to surrender in the face of recalcitrant data is the assumption that there must be predominantly *biological* reasons for inferior intellectual achievement in women.

Some have found this situation reminiscent of nineteenth-century craniometry's well-known attempt to explain inferior female intelligence by appealing to brain size. This is a case also discussed by Fausto-Sterling. The 'bigger is better' hypothesis foundered on the elephant problem (if absolute size were the true measure of intelligence, elephants should be smarter than people). So it was suggested that the true measure of intelligence lay in the proportion of brain mass to body mass; but this proportion favored women, and so the hypothesis was quickly rejected. The proposal that greater intelligence is linked to a lower ratio of facial bones to cranial bones ran afoul of the 'bird problem.' So it was suggested that the frontal lobes are the seat of the intellect, and men have bigger frontal lobes; the parietal lobes are larger in women. This hypothesis was surrendered when newer research pointed to the parietal lobes as the seat of the intellect. So the data were re-evaluated to show that *really* women have smaller parietal lobes . . . and so the saga continued. The one component of the theoretical network that scientists were unwilling to give up in the face of apparent falsification was the underlying assumption that women are *biologically* determined to be less intelligent than men. It is no wonder that feminist critics find the same pattern reinstated in current debates about gender and mathematical ability.

In the preceding cases, appeal has been made to such standard philosophical theses as the theory-ladenness of observation, the underdetermination thesis, and the Duhem-Quine thesis in order to suggest how gender ideology could permeate the biological sciences even on fairly standard accounts of theory appraisal. In these cases, we might want to say that external values have been imported into science; but the values are *implicit* in these cases and often exposed only in light of a rival hypothesis embedding conflicting values. The situation is different in the last set of cases in this rapid review of the literature. In the medical sciences, values or norms are often quite explicit. When one has to judge who is healthy and who is diseased, what body types are desirable and which not, the concepts involved are explicitly normative as well as descriptive. This opens the door for types of gender bias other than those discussed above. In one type, different ideals are set for male and female; these ideals are said to be 'complementary' but really only the male is seen as fully human. Another type of bias occurs when a single norm is adopted for both males and females, but is in actuality a male rather than human norm.

A nice historical example of the complementarity problem is developed in Londa Schiebinger's excellent book, *The Mind Has No Sex? Women in the Origins of Modern Science*.⁵ Schiebinger documents the changes that occurred in representations of male and female anatomy as a concerted effort was made in the eighteenth century to ground gender differences in anatomy. If differences between masculinity and femininity could be located in the *bones* of the organism, in its infrastructure, then there would be a modern scientific account of difference, and it would no longer be necessary to rely on the old heat models of Aristotle and Galen to do the job.

Prior to this time, male and female skeletons had been portrayed as similar; they were not sexualized. Sometimes the sex of the skeleton was not identified; sometimes the front view was represented as male, the back view as female. But all this changed in the years between 1730 and 1790.

The materialism of the age led anatomists to look first to the skeleton, as the hardest part of the body, to provide a "ground plan" for the body and give a "certain and natural" direction to the muscles and other parts of the body attached to it. If sex differences could be found in the skeleton, then sexual identity would no longer depend on differing degrees of heat (as the ancients had thought), nor would it be a matter of sex organs appended to a neutral human body (as Vesalius had thought). Instead, sexuality would be seen as penetrating every muscle, vein, and organ attached to and moulded by the skeleton.⁶

The male and female ideals that emerged were very different from one another. The male skeleton was typified by a big head and strong shoulders; its animal analogue was the horse, which sometimes appeared

in the background of male skeletal drawings. The female skeleton had a large pelvis, a long elegant neck, and a smallish head. She had much in common with the ostrich who sometimes decorated her portrait. Those skeletons which approximated most closely to cultural ideals of masculinity and femininity were favored for just that reason over drawings that were in some sense more accurate.

It is worth noting that one way the largeness of the female pelvis and smallness of the head are emphasized is by depicting a very narrow rib cage. Fausto-Sterling points out that there may have been more than just the power of ideology at work here. It may be that some of the corpses on which the drawings were modeled had their rib cages compressed by long-term use of the corset. This reminds us again that Ruth Hubbard⁷ and others are correct when they argue that it is wrong to think of the body as a purely biological infrastructure onto which the socio-cultural crud of gender accretes. Although the distinction between 'sex' as biological and 'gender' as socially assigned has in many respects served feminist theorizing well, it has sometimes led to the mistaken assumption that all biological attributes are given in some absolute sense. Sex as well as gender is socially constructed, at least in part. Such 'physical givens' as height, bone density, and musculature are to a large extent determined by cultural practice.

The skeletal case is one in which the male and female norms are said to be complementary, but the male is treated as more fully human. In other cases, there is allegedly a single human norm, but on closer inspection it turns out to be masculine. It has been suggested that the treatment of menstruation, pregnancy, and childbirth as diseases or medical emergencies may be traced to the fact that these are not things that happen to the ideal healthy human being who is, of course, male. The ideal healthy lab rat is also male. His body, his hormones, and his behaviors define the norm; so he is used in experiments. Female hormones and their effects are just *nuisance variables* that muck up the works, preventing experimenters from getting at the pure, clean, stripped-down essence of rathood as instantiated by the male model. Insofar as the female of the species is truly a rat (or truly a human being), she is covered by the research on males. Insofar as she is not included in that research, it is because she is not an archetypal member of her own species. The dangerous effects of such research procedures, especially in the biomedical sciences, are just now being documented. For far too long, the assumption underlying these experimental designs (that males are the norm) simply went unchallenged.

Elisabeth Lloyd is writing a book called *All About Eve* on the development of the female orgasm, and it includes a lovely example of a male norm masquerading as human. Various sociobiologists, when advancing theories about the evolutionary origins of the female orgasm, have cited detailed statistics about the nature, length, frequency, and repeatability of orgasm in order to support their origin stories. When tracking down their

footnotes, Lloyd discovered that these statements—which were being used to explain the origins of the *female* orgasm—were in fact based on data about *male* orgasms. The sleight of hand was typically accomplished by referring to the male subjects as ‘individuals’ rather than males!

II | Varieties of Feminist Critique

Case studies such as those canvassed above are interesting in their own right, but they leave open the question of what we are to make of them. Two contexts in which this question arises interest me particularly.

- 1 In the feminist literature the question that has been foremost in the last few years is whether these case studies are examples of ‘bad science’ or whether, on the contrary, they show that science is intrinsically and irredeemably androcentric.
- 2 In philosophy of science the question too often has been: what does this have to do with philosophy of science?

The two questions are related, and I should like to tackle them together. With respect to the first, Sandra Harding’s tripartite taxonomy of feminist epistemologies has been extremely influential.⁸ In order to deal with the bewildering diversity of feminist critiques of science, Harding proposes dividing them into three categories: feminist empiricism, standpoint epistemologies, and feminist postmodernism.

‘Feminist empiricism’ diagnoses failures such as those sketched above as failures of science to live up to its own ideals. Androcentric bias has gotten in the way of rigorous application of scientific method; but if the canons of science had been adhered to faithfully, episodes such as those above could have been avoided. For feminist empiricism, the standpoint of the knower is epistemically irrelevant; any bias originating from that standpoint will be eliminated by proper application of objective methods.

This assumption is denied by ‘standpoint epistemologists’ who argue that the credentials of the knowledge claim depend in part on the situation of the knower. Just as Hegel’s slave could know more than the master, so women (or feminists) may enjoy an epistemic advantage over men (or non-feminists). A science based upon the standpoint of women would be an improvement over current science, according to standpoint epistemology. In this sense it is still a ‘successor science’ project, since its aim is to produce a *better* (epistemically superior) account of the world. A number of problems have been pointed out with this approach, but the most damaging criticism has been the insistence that there is no single feminist standpoint. Just as the standpoint of women differs from that of men, so also the standpoint of poor women differs from that of rich women, the

standpoint of black women from that of white women, the standpoint of lesbians from that of heterosexual women, and so on. On what grounds could one of these be privileged over the other as a standpoint from which to describe the world?

This fracturing of identities and hence of standpoints has led some theorists to embrace what Harding calls 'feminist postmodernism' by giving up altogether the endeavor to become more and more objective and by accepting the existence of an irreducible plurality of alternative narratives about the way the world is. The notion of a scientific method that might allow us to transcend the constraints of culture, time, and place is repudiated once and for all by feminist postmodernists. Transcendental criteria for rationality and objectivity are dismissed as products of a masculine mythology, and the 'successor science' project is abandoned.

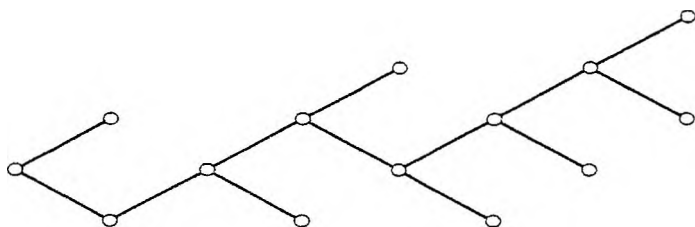
Although Harding's taxonomy has been very helpful in facilitating analyses of the diverse philosophical commitments of feminist critics, I fear that it also tends to obscure a promising possibility, one that takes into account the ways in which social structures (like the institution of gender) affect the very content of science without surrendering altogether the ideal of rational theory choice. In the following section, such a position is sketched.

III | Science and Shared Social Values

Traditionally, philosophy of science has been quite willing to grant that social and psychological factors (including perhaps gender) play a role in science; but that role has been a strictly delimited one, contained entirely within the so-called context of discovery, or alternatively within those episodes called 'bad science' in which the canons of rationality were clearly violated in favor of other interests. (The Lysenko Affair is a standard example here.) In the context of discovery or theory generation, says the traditional story, anything goes: the source of one's hypotheses is epistemically irrelevant; all that matters is the context of justification. If you arrived at your hypothesis by reading tea leaves, it doesn't matter so long as the hypothesis is confirmed or corroborated in the context of justification. You test the hypothesis in the tribunal of nature and if it holds up, then you're justified in holding on to it—whatever its origins. The idea here is that the canons of scientific theory choice supply a sort of filter which removes social, psychological, and political contaminants as a hypothesis passes from one context to the next.

This view made a certain amount of sense in the first half of this century when models of theory evaluation held that hypotheses were compared directly to nature. But this account, which shears the context of discovery or theory generation of all epistemic significance, makes no

sense at all given current models of scientific rationality that view theory choice as irreducibly *comparative*. That is, we now recognize that one does not actually compare the test hypothesis to nature directly in the hope of getting a 'yes' or 'no' ('true' or 'false') answer; nor does one compare it to all logically possible rival hypotheses. We can only compare a hypothesis to its extant rivals—that is, to other hypotheses which have actually been articulated to account for phenomena in the same domain and developed to the point of being testable. So the picture underlying current debates regarding theory choice looks something like this:



Each of the nodes is meant to represent a decision point at which the scientist must choose among alternative rivals. Methodological objectivists argue that so long as the proper machinery of theory assessment is brought to bear at each of the nodes, the rationality of science is preserved. How the nodes were generated in the first place is irrelevant, so long as the right decisions are made at each juncture. There may be interesting sociological stories to tell about the generation of the various alternative hypotheses, but sociological influences are effectively screened from affecting the content of science by the decision procedure operating at the nodes. This procedure will tell us which theory is preferable to its extant rivals on purely objective grounds.

My point, however, is that even if we grant for the sake of argument that scientific method is itself free of contamination by non-cognitive factors and that the decision procedure operates perfectly at the nodes, nothing in this procedure will insulate the content of science from sociological influences *once we grant that these influences do affect theory generation*. If our choice among rivals is irreducibly comparative, as it is on this model, then scientific methodology cannot guarantee (even on the most optimistic scenario) that the preferred theory is true—only that it is epistemically superior to the other *actually available* contenders. But if all these contenders have been affected by sociological factors, nothing in the appraisal machinery will completely 'purify' the successful theory.

Suppose, for the sake of example, that the graph represents the history of theories about female behavior. These theories may in many respects be quite different from one another; but if they have all been generated by males operating in a deeply sexist culture, then it is likely that all will

be contaminated by sexism. Non-sexist rivals will never even be generated. Hence the theory which is selected by the canons of scientific appraisal will simply be the best of the sexist rivals; and the very *content* of science will be sexist, no matter how rigorously we apply objective standards of assessment in the context of justification. In fact, the best of the sexist theories will emerge more and more highly *confirmed* after successive tests.

So, if my account is right, it doesn't necessarily follow that the presence of androcentrism and sexism in science makes rational theory choice impossible, but it *does* follow that scientific method *by itself* as currently understood cannot be counted upon to eliminate sexist or androcentric bias from science. Note that methodological rationalists can still have (approximately) monotonic progress. Every choice among alternatives may be a rational choice. Science can (in principle) get better and better. But this in no way guarantees that the content of science is insulated against social influences. Once you grant that social factors may influence the context of theory generation, then you *have* to admit that they may also influence the content of science. You can't just give theory generation to the social scientists and expect to exclude them at some later date through the rigorous application of epistemic virtue. That is akin to closing the barn door after the horses have escaped.

Let me make the same point in a different way. One of David Bloor's⁹ favorite arguments in support of social influences on theory content is based on the well-known underdetermination thesis. This is, of course, the claim that the data cannot pick out a single theory which uniquely accounts for them. There are, in principle, an infinite number of rival contenders that could do the same job. So, Bloor argues that if the data aren't completely determining our theory choices, then something else must be doing the job—and, of course, his favorite candidates for that job are sociological in character.

Larry Laudan's reply¹⁰ is that there is an unfortunate tendency in the recent literature to overestimate underdetermination. Underdetermination, he says, would be a problem if we were actually faced with an infinitude or even a pair of empirically adequate rival theories. But, as a matter of fact, he says, we never encounter such an embarrassment of riches. We're lucky if we get even *two* rivals that are credible contenders for theory acceptance. We're certainly never faced with more than a small handful of competing alternatives. And we can always find (at least in the passage of time) good cognitive reasons for preferring one of these to the others. So, he concludes, although the underdetermination thesis may pose nice problems in principle, these never figure into *actual* theory choice. This is, of course, another way of stating his claim that theory choice is irreducibly comparative in nature—that our choices will always be among a finite class of extant alternatives, not among an infinitude of in-principle rivals. Since there will always be good reasons for preferring one of the

extant rivals to the others, he claims that Bloor's invocation of social determinants is effectively undercut.

Notice, however, that there is an important sense in which this argument strategy simply begs the question. We can still ask why just *this* class of contenders was generated, given that others were equally compatible with the data. To say that once the rivals are fully articulated, our choice among them can be rational is to leave untouched the prior question of how our options came to be determined in the particular ways that they are. As long as Laudan concedes (as he does) that non-cognitive factors play a role in the posing of questions, the weighting of problems, and the initial articulation of theory, he cannot be sure that these factors will be eliminated in the context of justification. I stress once again that it is his attempt to maintain the *conjunction* of two views that gets him into trouble. The first of these views is that the context of discovery is normatively insignificant; the second is that theory appraisal is irreducibly comparative in nature. Once the second claim is made we *must* grant that factors affecting theory generation acquire normative significance.

The argument here is *not* that we should abolish the distinction between contexts of discovery and justification, but that we must recognize that on a comparative model factors that influence theory development and theory generation must necessarily influence our confirmation practices and hence the very content of science.

It is important to stress here that this argument about confirmation practices applies not only to test hypotheses but also to the auxiliary assumptions that jointly constitute the relevant background theory. How a particular piece of evidence bears on a hypothesis depends in large measure upon the collateral assumptions that come into play. It is here that the relationship between biology and the social sciences is particularly interesting because the traffic between the two is largely at this level. This is illustrated in some of the examples I cited earlier in this paper. For instance, in the man-the-hunter example, the relevant auxiliary assumptions are imported from evolutionary biology. In particular, it is the assumption that it is the male struggle for survival that drives the human evolutionary process ~~that dictates~~ in large measure what should count as evidence and how it should be interpreted. Conversely, in the Sleeping Beauty/Prince Charming model of the egg-sperm interaction, the biological data are informed by sociological assumptions about appropriate male and female roles. Donna Haraway's work in primatology¹¹ provides nice examples of how experimental design is influenced by background assumptions. She has traced the development of primatology since 1900, showing how political principles of hierarchy and male dominance have been embedded in that science, re-enforcing a theory of primate social organization in which a large, aggressive male is portrayed as defending a hierarchically organized troop and territory, enjoying first choice in food,

sex, and grooming, and deciding troop movements. Consequently, when Carpenter undertook his highly acclaimed work on rhesus monkeys, he removed dominant males to test his organizing hypothesis about the source of social order but undertook no control study in which other members of the group were removed. We can't control for every possible variable in our experimental designs; so which we take into account depends on what our background theory tells us may be relevant. If the components of that background theory are never called into question or challenged by a serious rival, our experimental practices will continue to embody potentially problematic assumptions.

The claim here is *not* that the traffic in auxiliary assumptions makes a pernicious form of holism inevitable or that these auxiliaries are not themselves (potentially) testable,¹² but that they provide points at which gender biases from one discipline are easily transported into another. Furthermore, because of the pervasiveness of gender ideology in our culture, these assumptions generally are not called into question and are sometimes not even noticed. It is usually the case that they come to light only in the presence of a rival hypothesis.

The argument here is not restricted to hypothetico-deductive forms of confirmation and cannot be evaded by an appeal to Clark Glymour's 'bootstrapping' model.¹³ Bootstrap confirmation does not make background assumptions dispensable but explicitly recognizes their crucial role: 'Hypotheses are not tested by themselves but only in relation to their fellows within the theory. Confirmation is a three-place relation, not a two-place relation. Large parts of the theory may be invoked in confirming, from given evidence, any of its hypotheses.'¹⁴

I have been arguing all along that *even if we grant that the standards of theory assessment are free of contamination by non-cognitive factors*, nonetheless, non-cognitive values may permeate the very content of science. Stating the thesis in this way seemed useful because it avoided the messy controversy regarding the culture-bound evolution of scientific method itself. Even *granting* the transcendence of method, in other words, the scientific product could itself be radically culture-bound.

I should mention in bringing this line of argument to a close, however, that what has been granted for the sake of argument is probably not plausible in the final analysis. Scientific method itself is developed and articulated by culture-bound individuals and so the arguments which applied at the object level of theory content will likely apply at the meta-level of theory evaluation as well. Although we may have *good reasons* for making certain methodological changes, (e.g., for moving from single blind to double blind experiments), our methodological choices will be limited by the range of alternatives already actualized.

Finally, I should touch very briefly on the implications of the preceding argument regarding the scope of models of rationality and its implications for science policy.

These appear to be the two alternatives: (1) We could simply acknowledge the reduced scope of models of rationality and make more modest claims for the objectivity of science; or (2) We could attempt to enlarge our model of rationality so that it takes into account the context of theory generation. That is, if we acknowledge that the context of theory generation has normative significance, then we may want to alter science policy in the light of a new normative account of theory generation.

Once we recognize that the content of science is affected by the social arrangements that govern its practice and production, then those social arrangements acquire *epistemic* significance as do the affirmative action programs and other interventions undertaken to alter those social arrangements. Any adequate philosophy of science will have to take this into account.

IV | Reviewing the Situation¹⁵

How does the account sketched above fit into Harding's taxonomy of feminist critiques of science? Clearly it shares much in common with so-called 'feminist-empiricism' insofar as it is a successor science project that aims at ever-increasing objectivity and rationality through the use of established scientific methods. It parts company with feminist empiricism, however, in at least two important respects. First, it recognizes that current methodologies simply do not take into account the epistemic significance of the social arrangements that govern the activities scientists undertake and the products they produce. Any adequate methodology will have to control for the biases introduced by these social arrangements just as it has to control for other sources of bias. (It has become fashionable recently to eschew talk of 'bias' on the ground that such talk implies the possibility of science that is entirely free of bias. I don't think the implication holds, and so I continue to speak of gender bias. We aim to eliminate other forms of partiality without thinking we'll ever be entirely successful; the same regulative ideal seems perfectly serviceable in discussions of androcentrism.)

Second, the feminist empiricism described by Harding does not appear to challenge the assumption in much traditional methodology that the rationality of the scientific community is just individual rationality writ large, a simple summation of individual rationalities. In the account sketched above, it is the rationality of the scientific community that is enhanced by inclusion of diverse strategies at the individual level. The kinds of bias discussed above can be systematically addressed only at the community level; no adequate program of *individual* rehabilitation could be prescribed in advance. Only the inclusion of diverse standpoints will bring about the conditions under which change is possible.

Is the current proposal then a kind of standpoint epistemology? Not precisely: epistemic privilege on this analysis does not attach to the individual woman (or feminist) but to the community that includes her standpoint along with others. The individual standpoints on this account are starting points. Furthermore, it is important to stress that on this analysis nothing depends on women having a different psychological make-up from men or different 'ways of knowing.' The distinctive mark of the work of the feminist critics cited above is not that it is holistic, intuitive, subjective, emotional, nurturant, or non-linear. Instead, what gives it focus and distinction is the fact that it is informed by a social and political viewpoint different from that which has dominated science and science studies.

Couldn't men have done exactly the same work? Yes, it is logically possible. But the connection here is not about necessary or sufficient conditions. It is about contingencies: about causal factors that operate, not from a God's eye point of view nor in the infinite long run, but here and now. It is not a logical necessity but also no accident that the advent of certain scientific hypotheses coincided with increased political power for women and increased representation of women in the academy and scientific communities.

Does the position advanced here have any affinities with feminist postmodernism? The overlap is minimal but not non-existent. My position is perfectly compatible with the rejection of metaphysical realism (perhaps that is even required) but not with the wholesale rejection of objectivity and rationality. The important point is that these two (metaphysical realism and objectivity) are separable, a point too often obscured in the postmodern literature. (Indeed, one often gets the impression from postmodern accounts that the logical positivists were metaphysical realists.)

I find feminist postmodernism unattractive for the usual reasons. I believe that feminist theories in science are *superior to* (cognitively preferable to) their sexist rivals, not simply that they provide alternative narratives. And I believe that postmodernism with its emphasis on fractured identities as well as on epistemic relativism provides no adequate basis for the political action feminism requires. There is much of value, however, in postmodernism's emphasis on the requirement of *local* problem-solving. Gender bias manifests itself in different ways in different sciences. There is no single 'feminist method' that will reveal and eliminate that bias. There is no 'feminist paradigm' that can be imposed from above and no reason to believe (as many postmoderns appear to believe) that gender bias in physics, for example, will be of the same kind or degree as that in biology. Real change in science will occur only when specific rival theories are developed by scientists who have both a thorough grounding in their own disciplines and a commitment to questioning biases introduced by social arrangements of science.

I believe, therefore, that it is possible to do justice to the range and

depth of gender bias in the biological sciences without sacrificing altogether the traditional ideals of objectivity and rationality; but doing so will require that we take into account the social structure of science. Case studies of the sort summarized in the first part of this paper show the necessity of coming to grips with the ways in which social factors can influence the development of science, and they demonstrate the extent to which some standard philosophical tools can partially illuminate the origins and diversity of ideological biases in science. These tools, however, are inadequate to the task at hand so long as they are embedded within an outmoded and indefensible conception of the scientific process that limits the influence of social factors to the context of discovery. Mainstream philosophy of science continues to ignore feminist critiques of science at its own peril.¹⁶

Notes

1. The Biology and Gender Study Group, 'Importance of Feminist Critiques for Contemporary Cell Biology,' in *Feminism and Science*, Nancy Tuana, ed. (Bloomington: Indiana University Press 1989) 172-87.
2. Charles Darwin, *The Descent of Man* (1871). Cited by Ruth Hubbard in 'Have Only Men Evolved?' in *Biological Woman: The Convenient Myth*, Ruth Hubbard, Mary Sue Henifin, and Barbara Fried, eds. (Cambridge, MA: Schenkman 1982), 17-45.
3. Helen Longino and Ruth Doell, 'Body, Bias, and Behavior: A Comparative Analysis of Reasoning in Two Areas of Biological Science,' *Signs* 9 (1983), 206-27.
4. Anne Fausto-Sterling, *Myths of Gender: Biological Theories About Women and Men* (New York: Basic Books 1985).
5. Londa Schiebinger, *The Mind Has No Sex? Women in the Origins of Modern Science* (Cambridge, MA: Harvard University Press 1989) 189-213.
6. *Ibid.*, 191.
7. See *The Politics of Women's Biology* (New Brunswick: Rutgers University Press 1990).
8. Sandra Harding, *The Science Question in Feminism* (Ithaca, NY: Cornell University Press 1986). See also S. Harding, *Whose Science? Whose Knowledge? Thinking From Women's Lives* (Ithaca, NY: Cornell University Press 1991).
9. David Bloor, *Knowledge and Social Imagery* (London: Routledge and Kegan Paul 1976).
10. Larry Laudan, 'The Pseudo-Science of Science?' *Philosophy of the Social Sciences* 11 (1981), 195-97.
11. See, for example, 'Primateology Is Politics by Other Means,' *PSA* 1984, vol. 2 (East Lansing, MI: Philosophy of Science Association 1985).

12. See Helen Longino, *Science as Social Knowledge* (Princeton: Princeton University Press 1990) [a portion of which is excerpted above, pp. 170–91].

13. Clark Glymour, *Theory and Evidence* (Princeton: Princeton University Press 1980).

14. *Ibid.*, 151.

15. The argument in this section was also sketched in my essay, 'Birth of a New Physics or Death of Nature?' in *Women and Reason*, E. Harvey and K. Okruhlik, eds. (Ann Arbor, MI: University of Michigan Press 1991).

16. Work on this project was supported by a grant from the Social Science and Humanities Research Council for which I am grateful. I would also like to thank J. R. Brown and Alison Wylie for useful discussions during the project's early days and my sister Peggy Okruhlik for making its completion possible. This paper results in part from amalgamating two earlier typescripts that circulated widely. The first was called 'A Locus of Values in Science' and dates from 1984; the second, 'Gender Ideology and Science,' was first drafted in 1988.

- 2.1 Values in Science 211
 - Cognitive Values* 211
 - Contextual Values* 212
 - The Value-Neutrality Thesis and Its Critics* 213
- 2.2 Kuhn's Analysis of Science 215
 - Paradigms and Normal Science* 215
 - Scientific Revolutions* 217
- 2.3 Six Kuhnian Arguments for Relativism 219
 - Theory-Ladenness of Observation* 220
 - Meaning Variance* 222
 - Problem Weighting* 224
 - Shifting Standards* 225
 - The Ambiguity of Shared Standards* 225
 - The Collective Inconsistency of Rules* 226
- 2.4 Kuhn's Second Thoughts about the Rationality of Scientific Revolutions 227
 - Shared Epistemic Values* 228
- 2.5 McMullin's Criticisms of Kuhn 230
 - Shared Values* 231
 - The Justification of Values* 232
 - Rationality and Realism* 233
- 2.6 Laudan's Criticisms of Kuhn 235
 - The Hierarchical Model of Scientific Rationality* 235
 - The Reticulational Model of Scientific Rationality* 236
 - The Flaws in Kuhn's Arguments* 237
- 2.7 Longino on Contextualism and Objectivity 239
 - The Logical Positivist versus the Contextualist Account of Scientific Method* 240
 - The Contextualist Analysis of Evidence* 241
 - Criteria for Objectivity* 242
- 2.8 Okruhlik on the Feminist Critique of Science 245
 - Varieties of Feminist Critique* 246
 - Okruhlik's Proposal* 248
- 2.9 Summary 249

2 | COMMENTARY

2.1 | Values in Science

Contrary to what you might expect, none of the readings in chapter 2 concern moral values. There has been an unfortunate tendency, especially in public debates about education, to assume that all values are moral values. But there are many different kinds of values that people approve of and wish to promote. They range from aesthetic to religious, from cognitive to noncognitive. Moral values—ideas about what things are morally good and which actions morally right—are just one sort of value among many. Of course, moral values are relevant to the practice of science. Scientists are people, after all, and, as such, their behavior should conform to the norms of morality. For example, there are moral constraints on what kinds of experiments can be performed (especially those involving human subjects) and how they should be conducted; there are rules concerning the use of other scientists' ideas and prohibitions against faking or fudging experimental data; there can also be serious moral concerns about the ways one's research might be used.¹ Clearly, then, there is a moral dimension to science just as there is with any human activity. But when the authors in this chapter talk about values in science, they are not referring to moral values; instead, their focus is on either cognitive or contextual values.

COGNITIVE VALUES

Cognitive values (sometimes called *epistemic values*) have to do with cognition or knowledge. Now, one might think that the paramount epistemic value in science is truth: scientists have as their principal aim the acceptance of true claims about the natural world (and the rejection of false ones). But on closer inspection, this assumption proves flawed. For one thing, truth by itself cannot be sufficient as a characterization of the goal of science. Why not? Because so many of the true statements we could make about the natural world have little or no scientific value. Imagine, for example, that a biologist wants to increase our store of scientific knowledge by counting the precise number of hairs on individual dogs at various times on various days, not to test a theory or experiment with a drug to prevent hair loss but simply to know the canine hair-count for its own sake. Even if the information that the biologist collects is true, it has negligible scientific value. Scientists are interested not merely in discovering truths about the world, but in discovering interesting truths, usually in the form of general theories and laws with predictive power. These

criteria of scientific excellence—generality and predictive power—and many others besides (such as explanatory power and simplicity) are among the cognitive values of science. They are not the same as truth, and despite their relation to truth, they do not seem to be reducible to truth in any obvious way.

Cognitive values are, in an important way, partly constitutive of what science is. They help to define the nature of science and to distinguish it from other truth-seeking enterprises, such as genealogy, logic, and theology (to name but three). They underlie many of our judgments about what constitutes acceptable scientific method and practice. For this reason, philosopher of science Helen Longino refers to them as *constitutive values*. In the readings for chapter 2, there is much discussion of whether the existence of several different, sometimes competing, cognitive (constitutive) values compromises the objectivity of science or undermines the rationality of scientific change.

A second problem with taking truth as the sole epistemic value is that scientists cannot tell whether a theory is true simply by inspecting it; they need to perform experiments and make observations. Interesting scientific theories involve empirical generalizations of universal scope and are often about entities and processes that cannot be directly observed no matter how sharp one's vision. Thus in order to reach judgments about the truth of a theory, scientists must rely on other criteria that, they hope, have some connection with truth. For example, commonly a theory is believed more likely to be true if it gives rise to correct predictions. (See chapters 4 and 5 for a detailed discussion of the relationship between prediction and confirmation.) Similarly, but more controversially, a theory is believed more likely to be true if it explains a wide variety of phenomena. More controversially still, a theory is believed more likely to be true if, other things being equal, it is simpler than its rivals. The moral is clear: even if truth is the primary cognitive value, we need other cognitive values to help us decide which of our theories are true. When choosing between competing theories, various cognitive values, such as predictive power, explanatory scope, and simplicity, may not always point in the same direction, so reasonable disagreements might arise about whether (or to what extent) these values are indicators of truth. Thus, even within the realm of cognitive values, debates about the relative importance of particular values and their merits as indicators of truth are bound to arise. So even if we accept that truth is the primary virtue of a scientific theory, value debates within science are inevitable.

CONTEXTUAL VALUES

What are contextual values? Roughly speaking, they are norms, preferences, beliefs, and interests that are unrelated to the cognitive aims of science. For example, in the 1930s the official German scientific estab-

lishment became so infected with Nazi ideology that it rejected Einstein's theory of relativity on the grounds that its author was Jewish. Limiting one's teaching and research solely to theories whose authors are Aryan would be to adopt an obnoxious (and stupid) contextual value in the practice of science. More subtle examples of contextual values include the preference for theories (especially in the social and biological sciences) that conform to widespread, unexamined, and usually false beliefs about gender and race: in short, racial and sexual bias.

Helen Longino uses the term *contextual values* because these values vary with time and across cultures (and even from one scientist to another). Contextual values show a much greater degree of dependence on society and culture than do their cognitive counterparts because they are not tied in any close and essential way to the fundamental aims of science, such as the search for truth. Rejecting the constitutive (cognitive) values of science is impossible without ceasing to do science, but scientists can (and do) have different sets of contextual values that can (and do) affect the way they do science.

THE VALUE-NEUTRALITY THESIS AND ITS CRITICS

Until quite recently, most philosophers of science accepted the standard view that, insofar as science is value laden, the values in question are exclusively cognitive ones. Contextual values have no place in science; or, at least, *should* have no place. Because contextual values are unrelated to the fundamental goals of science, they should play no role in scientific decision making. In that sense, science should be value-free (i.e., free of contextual values, not free of values of any kind). This view is also called the *value-neutrality thesis*. According to the value-neutrality thesis, when contextual values intrude into science, their influence is invariably pernicious. "Good" science is objective and thus free from subjective values, whether they are the subjective preferences of a single individual or the cultural biases of an entire society. Contextual values distort the judgment of individual scientists and diminish scientific objectivity. Although it is an empirical issue (and thus not one that philosophers have any particular competence to address), proponents of the standard view assume that episodes of contextually infected "bad" science have been historically rare. Indeed, those holding the standard view assume that the constitutive values of science and the institutional mechanisms based on them are a prophylaxis against such infection and an effective cure for it when it occasionally occurs.

As we shall see in our later discussion, Helen Longino rejects several key aspects of the standard view and the value-neutrality thesis. She defends the view—referred to as *contextualism*—that all scientific judgments depend, to a greater or lesser degree, on contextual values. But, she argues, this value-ladenness of science is not incompatible with scientific objec-

tivity. Longino focuses primarily on the appraisal of theories and hypotheses after they have been formulated and examines the process by which scientists decide which are confirmed by the available evidence. She argues that a proper analysis of the relation between evidence and theory shows that all judgments about when and to what extent a piece of evidence confirms a theory depend on background beliefs. Since these background beliefs can vary widely, different scientists can reach different judgments about the significance of an observation or experiment. Longino calls this the *contextualist analysis of evidence*.

If the contextualist analysis of evidence is correct and all judgments about confirmation and evidence are relative to the subjective beliefs of the individual researcher, how can science be objective? The answer, Longino thinks, is to realize that scientific knowledge is the product of a community of scientists. Scientific knowledge is, she claims, social knowledge. As such, it emerges through a process of criticism (such as peer review) based on shared community standards (not individual ones). The social nature of objectivity does not guarantee that science will be free from all subjective preference, but if certain criteria are met (she lists four: recognized avenues for criticisms, shared standards, community response, and equality of intellectual authority), to that extent science will be objective. Thus, in several respects, Longino challenges the value-neutrality thesis: she denies that scientists are free from contextual values when they decide which theories are confirmed by the evidence; she rejects the contention that scientists ought to be free from all contextual values as an unrealizable ideal resulting from a faulty understanding of the relationship between theory and evidence; and she insists that the neutrality thesis is mistaken in assuming that contextual values necessarily compromise the objectivity of science.

Okruhlik's article complements Longino's by describing several cases of sexist bias in modern biological research. Unlike some feminist critics of science (such as Sandra Harding), Okruhlik does not think that we should abandon rationality and objectivity as self-serving male myths. Rather, Okruhlik argues that once we understand how science works, especially in coming up with hypotheses in the first place, we can retain the notion of scientific rationality while recognizing that the scientific method cannot, by itself, be relied upon to eliminate sexist bias. One of Okruhlik's key arguments is that, even if the process of choosing among rival theories were fully objective and rational, if all the theories under consideration are sexist so, too, will be the theory selected as the best. Thus, even though theory appraisal may be free from bias, the contamination of theory generation by contextual factors will inevitably affect the content of science.

2.2 | Kuhn's Analysis of Science

PARADIGMS AND NORMAL SCIENCE

In *The Structure of Scientific Revolutions*, Kuhn portrays scientific disciplines as going through three stages: immature science, normal science, and revolutionary science. At the heart of Kuhn's characterization of each of these stages is his notion of a paradigm. Stage one occurs when an area of inquiry first attains the status of a genuine science. For Kuhn, this transition from immature science (or nonscience) to mature science is marked by the discipline acquiring its first paradigm. At least in *The Structure of Scientific Revolutions*, Kuhn portrays stage one as occurring only once in the history of a scientific field, and he refers to immature science ("something less than science") as being preparadigmatic.² To use terminology introduced in chapter 1 of this book, Kuhn is treating the possession of a paradigm as a *demarkation criterion*, something that essentially distinguishes genuine science from things that are not science (or which might merely pretend to be science).

Stage two is what Kuhn calls *normal science*. Normal science is the way science is usually done (except for those relatively rare periods of extraordinary or revolutionary science that Kuhn considers stage three). The hallmark of normal science is that scientists work under the aegis of a paradigm that they all accept. Typical activities of normal science are making more accurate measurements of constants, looking for entities and processes that the paradigm tells us must exist, extending the paradigm to new areas and types of phenomena, reconciling the paradigm with recalcitrant data, and removing the conceptual difficulties that afflict even the most successful paradigms. Kuhn refers to all these paradigm-based activities as *puzzle solving*. To call something a puzzle (whether it is a crossword puzzle, a jigsaw puzzle, or a scientific puzzle) is to imply that it has a solution. The failure of a scientist to solve a scientific puzzle reflects on the competence of the researcher, not on the soundness of the paradigm. Thus, although paradigms typically float in a sea of anomalies, normal scientific activity is directed towards removing those anomalies, not towards refuting the paradigm.

Kuhn portrays normal science as dogmatic: students are taught the paradigm without being encouraged to question it or seriously consider alternatives; debates about fundamentals are discouraged or suppressed; the main scientific activity—puzzle solving—takes for granted that the prevailing paradigm is basically correct. (For more on Kuhn's account of normal science, see his "Logic of Discovery or Psychology of Research?" in chapter 1 of this book and the section "Kuhn's Criticisms of Popper" in the commentary on that chapter.)

Although the concept of a paradigm is fundamental to Kuhn's analysis of science, he does not define it clearly and precisely in *The Structure of*

Scientific Revolutions. In his *Postscript*, Kuhn acknowledges his critics' complaint that he had used the term *paradigm* ambiguously. In response to this criticism, Kuhn introduces two new terms to stand for two of the most important senses of *paradigm*: *exemplars* and *disciplinary matrices*. Exemplars include the solutions to problems used in the education of scientists: typically these solutions appear at the end of chapters in science texts, in laboratory exercises, and on examinations. More advanced technical solutions are published in journals. Kuhn's point about these exemplars is that they are crucial to teaching students how to use theories to solve problems; merely memorizing a few equations is not enough. Students become scientists by recognizing and embracing the tacit knowledge implicit in the exemplars.

Kuhn uses the phrase *disciplinary matrix* to stand for his second sense of *paradigm*. Disciplinary matrices include exemplars, but they also contain, among their main components, symbolic generalizations, metaphysical commitments, heuristic models, and values. It is this second, more inclusive sense of *paradigm* that Kuhn employs in many of his more challenging and controversial claims about science. Henceforth, when we speak of paradigms, it will be in terms of Kuhn's notion of disciplinary matrices.

Kuhn makes several claims about normal science that have elicited much criticism. These claims are important because they underlie many of the assertions that Kuhn makes about scientific revolutions. Here we will simply list these claims with brief comments, reserving criticisms for later.

■ NORMAL SCIENCE IS BASED ON A PARADIGM.

The fundamental entity that drives normal science is not an individual theory but a paradigm. Paradigms are collections of diverse elements (ontological, epistemological, methodological, and axiological) that tell scientists what experiments to perform, which observations to make, how to modify their theories, how to make choices between competing theories and hypotheses, and so on.

■ NORMAL SCIENCE IS DOGMATIC.

Normal science is a relatively dogmatic affair, consisting largely of what Kuhn calls puzzle solving. In Kuhn's words, normal science is a "mopping-up operation," "an attempt to force nature into the preformed and relatively inflexible box that the paradigm supplies" (*The Structure of Scientific Revolutions*, 24).

■ NORMAL SCIENCE IS OBJECTIVELY PROGRESSIVE.

Kuhn equates progress with cumulativeness. Since (when things are going well) scientists accumulate solutions to scientific puzzles, normal science must be progressive. Thus, at least with respect to normal science, Kuhn

agrees that science is progressive in an objective sense. When we turn to scientific revolutions, things are radically different.

Because of their importance in providing philosophical support for Kuhn's claims about both normal and revolutionary science, we list the doctrines underlying Kuhn's arguments. Later, we shall examine these doctrines and the arguments based upon them in greater depth.

■ THE HOLISTIC CHARACTER OF PARADIGMS

Paradigms contain a variety of elements—from explicit claims about the kinds of entities and processes that populate the world to implicit commitments to various methodological values, from precisely defined techniques to loosely articulated heuristic models. Nonetheless, Kuhn presupposes that paradigms are integrated wholes. Because the elements of a paradigm are inextricably intertwined, scientists must accept or reject the paradigm as whole; they cannot accept some elements and reject others.

■ THE THEORY-LADENNESS OF OBSERVATION

The theories that scientists accept (and the paradigms in which they are embedded) significantly affect what scientists observe. In *The Structure of Scientific Revolutions*, Kuhn subscribes to the thesis of the theory-ladenness of observation, according to which what scientists observe is a function of what they believe.

■ THE THEORY-DEPENDENCE OF MEANING

The theories that scientists accept (and the paradigms in which they are embedded) significantly affect what scientists mean by the theoretical terms they employ. In particular, theoretical terms such as *planet*, *electron*, *oxygen*, *gene*, and *mass* derive their meaning, at least in part, from the theories in which they occur.

SCIENTIFIC REVOLUTIONS

In "The Nature and Necessity of Scientific Revolutions," Kuhn tells us that "scientific revolutions are here taken to be those non-cumulative developmental episodes in which an older paradigm is replaced in whole or in part by an incompatible new one" (86). In brief, scientific revolutions occur when scientists switch from one paradigm to another.

It is clear from *The Structure of Scientific Revolutions* that Kuhn does not intend simply to define scientific revolutions as noncumulative and leave it at that. Nor does his claim rest merely on history, on scientific revolutions past. Rather, while he frequently illustrates his claims by appealing to history, he also argues that scientific revolutions must be non-cumulative because of certain presumed truths about the nature of

paradigms, observation, and meaning.³ Thus, while Kuhn claims that his view of science accurately reflects the way science in fact works, he is not simply describing selected episodes from the history of science and generalizing from them. Like many of the philosophers of science that he criticizes, Kuhn, too, draws conclusions about the way science must develop on the basis of philosophical arguments. Let us begin with Kuhn's conclusions, and then examine his arguments.

Below is a summary of the six main claims that Kuhn makes about scientific revolutions in "The Nature and Necessity of Scientific Revolutions" and elsewhere in *The Structure of Scientific Revolutions*.

■ SCIENTIFIC REVOLUTIONS ARE PARADIGM SHIFTS.

Paradigms are as vital to our understanding of scientific revolutions as they are to our understanding of normal science. Scientific revolutions involve not merely the replacement of one theory by another but a wholesale shift from one network of scientific commitments, beliefs, and values to another.

■ SCIENTIFIC REVOLUTIONS ARE TOTAL.

Scientific revolutions are always total: "Though logical inclusiveness remains a permissible view of the relation between successive scientific theories, it is a historical implausibility" (*The Structure of Scientific Revolutions*, 98). Paradigms are mutually exclusive: one cannot consistently hold both; commitment to one paradigm logically precludes commitment to its rival.

■ SCIENTIFIC REVOLUTIONS ARE RELATIVELY SUDDEN AND UNSTRUCTURED EVENTS.

Kuhn compares scientific revolutions with gestalt switches, flashes of intuition, and religious conversions. Like some conversion experiences, scientific revolutions are preceded by a period of debate. Indeed, Kuhn argues that, given the dogmatism of normal science, a period of extraordinary or revolutionary science is required to relax the normal suppression of debate about fundamentals. Kuhn describes this period of debate as a crisis brought about by the accumulation of anomalies under the prevailing paradigm. But when the revolution finally occurs, it is rapid and unstructured. Hence, Kuhn's book appears to be mistitled, since, according to Kuhn, there is no structure of scientific revolutions. But the appearance of paradox is removed by distinguishing between revolutionary science (the structured process leading up to the paradigm switch) and a scientific revolution (the unstructured event of paradigm switch). We make much the same distinction between, say, the process of dying and the event of death.

■ EXTRAORDINARY (REVOLUTIONARY) SCIENCE IS NONDOGMATIC.

This is clear from the previous claim. Because normal science is so dogmatic, the process leading up to a paradigm switch must involve a much freer kind of scientific debate, in which dissent and questioning are tolerated. Otherwise, scientific revolutions would be impossible.

■ SCIENTIFIC REVOLUTIONS CANNOT BE DECIDED BY RATIONAL DEBATE.

Kuhn insists that scientific revolutions necessarily have an extrarational character. The choice between one paradigm and its rival cannot be decided using only the values and methods of normal science. Arguments in favor of a paradigm are bound to be circular, taking for granted doctrines that proponents of a rival paradigm reject. So paradigm acceptance via revolution is essentially a sociological, not a rational phenomenon—a matter of persuasion and conversion, not of rational argument. “As in political revolutions, so in paradigm choice—there is no standard higher than the assent of the relevant community” (*The Structure of Scientific Revolutions*, 94).

■ SCIENTIFIC REVOLUTIONS ARE NOT OBJECTIVELY PROGRESSIVE.

Time and again Kuhn proclaims that there is no genuine, objective progress through scientific revolutions; changes of paradigms cannot be said to bring us closer to the truth. Once a scientific revolution has occurred, the writers of the new textbooks and histories of science make the development of science *look* cumulative and linear, but that is simply window dressing. It has no more objective significance than did the histories of political development written by Marxist ideologues in the former Soviet Union. In science, as in politics, history is rewritten by the winners.

All six of Kuhn’s general claims about scientific revolutions have drawn fire from critics. But particular attention has been focused on the last two theses because they deny that science as a whole is either objectively rational or objectively progressive—a position that challenges the traditional image of science. According to Kuhn, judgments about scientific rationality and progress can only be made relative to paradigms. With good reason, then, Kuhn’s overall position is aptly described as a version of relativism.

2.3 | Six Kuhnian Arguments for Relativism

Kuhn has at least six different arguments for his conclusions about the lack of rationality and objective progress in scientific revolutions. These

concern the theory-ladenness of observation, meaning variance, problem weighting, shifting standards, the ambiguity of shared standards, and the collective inconsistency of rules. We shall briefly describe each of these arguments and criticize the first two; criticisms of the last four can be found in the discussions of McMullin and Laudan.

THEORY-LADENNESS OF OBSERVATION

In accounts of confirmation and disconfirmation, the fate of theories is often decided, at least in part, by observations. (For a more detailed consideration of confirmation, see chapters 4 and 5.) Typically, a theory makes a prediction concerning something we should be able to observe. Sometimes, the observation is just a matter of looking and seeing; more often, the observation requires the use of instruments or special apparatus, or the performance of an experiment. When two rival theories make incompatible predictions about the outcome of an experiment, we should be able to decide between them (at least in principle) by performing a crucial experiment. Indeed, the empirical component provided by observation usually is supposed to confer on science its objectivity and rationality: choices about scientific theories are considered objective and rational largely because acceptable theories must agree with observation. Science is considered progressive because current theories are in closer agreement with observations than were the theories that they replaced. Whether or not empirical evidence (such as that provided by observations) can ever determine theory choice is considered in chapter 3; for the moment the issue is whether observation can serve as a rational basis, however partial, for accepting or rejecting individual theories or for deciding between rival theories. Kuhn's argument regarding the theory-ladenness of observation denies that observation can play this role.

The key premise in Kuhn's argument is that observation is theory laden. What does this mean? Roughly, it means that what scientists observe depends on the theories they accept. In order for this to support Kuhn's denial that observation can serve as a rational basis for making decisions about theories, the thesis must imply that (1) no proponent of a scientific theory can ever observe anything contrary to that theory and that (2) scientists who accept rival theories can never observe the same thing.

When put in this stark fashion, neither (1) nor (2) seems at all plausible. Let us evaluate (2) by considering a simple example.⁴ Imagine that Kepler and Ptolemy are watching the sun climb over the horizon at sunrise. Kepler accepts the Copernican theory that the earth rotates on its axis once a day and revolves around the sun once a year. Ptolemy accepts the Ptolemaic theory that the earth is completely stationary and that it is the sun that makes both a daily rotation and an annual revolution around the earth. For Kepler, the earth is a planet and the sun is stationary; for Ptolemy, the sun is a planet and earth is at rest. Now, do Kepler and

Ptolemy see the same thing at dawn when they observe the rising sun? The proponent of the theory-ladenness thesis might argue thus: Kepler's sun is at rest and Ptolemy's sun is moving; it is logically impossible for the same object to be both at rest and in motion; therefore, Kepler's sun cannot be the same as Ptolemy's sun; therefore the object that Kepler observes (Kepler's sun) cannot be the same as the object that Ptolemy observes (Ptolemy's sun). But this argument is flawed because it fails to distinguish between the object that is observed (the sun) and the rival beliefs that the two observers have about it. Thus the argument commits the fallacy of equivocation. For example, the phrase *Kepler's sun* in the first premise means *the sun as Kepler believes it to be* (or, more strictly, *the thing that would answer to Kepler's beliefs if they were true*), but in the conclusion, *Kepler's sun* means *the object (the sun) that Kepler observes (regardless of Kepler's or anyone else's beliefs about it)*.

Another way of appreciating the implausibility of the thesis that proponents of rival theories observe different objects is to imagine that Ptolemy and Kepler are discussing the relative merits of their respective theories as dawn arrives. Suppose, as our literal construal of Kuhn's account requires, that the thing that Ptolemy sees is distinct from the thing that Kepler sees. (Never mind that this seems to require two heavenly bodies to be in exactly the same place.) Were Ptolemy to convert to Copernicanism (thanks to Kepler's arguments), we would be forced to accept the conclusion that there is suddenly one less star in the heavens. This would confer on Ptolemy the magical power of being able to change the world simply by changing his beliefs.

Sometimes the theory-ladenness thesis is expressed by saying that there is no (or cannot be any) theory-neutral observation language. Thus, even though it might be conceded that Kepler and Ptolemy are looking at (and hence observing) the same object, the words they use to describe their respective observations must have different meanings, depending on the different theories that they hold. Insofar as this new form of the argument involves considerations about language and meaning, it will be discussed in the next section, on the meaning-variance argument. For the moment, we will merely comment on the first implication of Kuhn's thesis, namely that no proponent of a scientific theory can ever observe anything contrary to that theory. Clearly, this is untenable. Even if it were true that every observation a scientist makes in the course of testing a theory involves terms from the theory that is being tested (and that the scientist accepts as true), it simply does not follow that the scientist will never find any observational evidence that disconfirms the theory. For we must distinguish between the concepts used to make a judgment and the assertions made using those concepts.⁵ For example, Kepler believed that all planets move in ellipses. Suppose that there is no theory-neutral observation language and that Kepler had to use the terms *planet* and *ellipse* in any statement reporting his observations for the statement to have an evidential

bearing on his theory. Even given these constraints Kepler could easily have refuted his own theory by reporting that the planet he observed did not move in an ellipse. (Just this happened with regard to some of Kepler's earlier hypotheses about the shape of planetary orbits before he hit on the idea of an ellipse). The necessity of using certain concepts in reporting one's observations in no way determines the truth or falsity of the judgments one makes using those concepts. Thus, while it is true that our theoretical beliefs and expectations sometimes influence our perceptions and so, to that extent, at least some of our perceptions are theory laden, we have been given no reason to think that what we perceive must always agree with what we believe.⁶

MEANING VARIANCE

The crucial premise in the meaning-variance argument is the thesis of the theory-dependence of meaning, according to which the theories scientists accept (and the paradigms associated with them) significantly affect what scientists mean by the theoretical terms they use. In order for this thesis to support Kuhn's denial of rationality and progress, it would have to entail that scientists committed to rival paradigms speak different languages. They may utter the same words, but the words have different meanings, so any logical comparison of their utterances is precluded, leaving adherents of rival theories simply talking past one another. Kuhn sometimes expresses this view by saying that rival paradigms and the terms they employ are *incommensurable*.

The term *incommensurable* literally means "having no common measure." It comes from geometry, where it was first used to express the fact that in an isosceles right-angled triangle whose side is one unit in length (a rational number), the hypotenuse (the square root of two—an irrational number) cannot be expressed as any ratio of rational numbers. The side and the hypotenuse lack a common measure. In a similar way, Kuhn claims that the proponents of rival paradigms lack a common language. In his later writings Kuhn has insisted that his view in no way precludes translation from one theoretical language to another. But even then, he claims, the translation of one theory into the idiom of another will not exactly preserve the meaning of the original. Thus, for example, Tycho Brahe could understand (and reject) the Copernican claim that the Earth is a planet moving around the sun. But, according to Kuhn, when Brahe uses the term *planet* in his rejection of Copernican theory, the word *planet* means something different for Brahe than it does for Copernicans such as Kepler. So, Kuhn concludes, there is no theory-neutral, paradigm-independent way of comparing the two theories.

Claims about meaning variance are notoriously hard to evaluate, mainly because there is no consensus about what meanings are in the first place. But this much is clear: whatever meanings are, if comparisons be-

tween theories are to involve logically valid arguments, then the terms such arguments employ must have the same, unequivocal meaning throughout. (Otherwise, the arguments in question would commit the fallacy of equivocation.) Thus, if meaning variance inevitably occurs whenever two theories differ, then valid logical comparisons between them would be impossible. In this way, some proponents of the meaning-variance thesis (such as Paul Feyerabend) have attacked models of reduction that require the old theory to be logically derivable from the theory that replaces it. (For more on Feyerabend, theory reduction, and meaning variance, see chapter 8.)

Many of Kuhn's critics have pointed out an apparent contradiction between Kuhn saying, on the one hand, that rival paradigms are incommensurable with respect to meanings and, on the other hand, that it is impossible to believe both paradigms at the same time. For, without commensurability of meaning, there cannot be logical incompatibility. Consider, for example, the word *planet*. Suppose that Brahe meant by *planet*, *satellite of the earth* and that Kepler meant by *planet*, *satellite of the sun*. Would they have been disagreeing when Brahe said, "All planets go around the earth," and Kepler said, "All planets go around the sun"? Since the word *planet* has different meanings in the two sentences, neither sentence contradicts the other. In short, there is no single proposition that one affirms and the other denies. Indeed, if the word *planet* were completely defined by Brahe and Kepler in the two ways indicated, then both Brahe and Kepler would have been saying something true but utterly trivial (Brahe: all planets [read: *satellites of the earth*] go around the earth; Kepler: all planets [read: *satellites of the sun*] go around the sun).

Clearly something has gone wrong in our telling of the story about Brahe and Kepler. It is absurd to suggest that they would have been asserting tautologies were they to have uttered the sentences attributed to them. The flaw in the story seems to lie in the assumption that there is nothing more to the meaning of the word *planet* than the role that the term plays in a particular theory. The word was used long before Brahe and Kepler to refer to astronomical bodies such as Venus, Mercury, and Jupiter. And it is this referential dimension of the term that enables Brahe and Kepler to make contrary claims about those bodies.

Expressed more formally, there are (at least) two aspects to the meaning of theoretical terms such as *planet*, *electron*, *gene*, and *oxygen*: their sense and their reference. The sense of a term in a particular theory is given partly by its use prior to its adoption by the theory (its antecedent meaning) and partly by its role in that theory. Only when a term is a neologism, a brand new term, will its sense be defined solely by the theory that employs it. The reference of a term is the set of things, processes, or events that the term picks out. In general, rival theories can make opposing claims about the world, despite some differences in the sense of some of the terms they use, *if* they make opposing claims about some of the things,

processes, or events that the terms pick out. This is only the barest beginnings of an account of the meaning of theoretical terms, but at least it enables us to see how theories can be genuine rivals despite the contribution that theoretical context makes to meaning. In short, radical incommensurability fails because there is more to the meaning of most terms than simply the sense defined implicitly by a particular theory (the "more" being what we have called antecedent meaning) and there is more to meaning than sense (the "more" in this case being reference).

The upshot of our discussion thus far is that neither of Kuhn's two arguments—neither the one based on the theory-ladenness of observation nor the one based on the paradigm-dependence of meaning—is strong enough to support his radical conclusions about the lack of rationality and progress in scientific revolutions. We now turn to four arguments that have proven more durable. These arguments, endorsed by philosophers of science favorable to Kuhn and criticized by McMullin and Laudan, are sketched, more or less explicitly, in the second reading from Kuhn, "Objectivity, Value Judgment, and Theory Choice." For convenience, we have adopted the labels that Laudan gives to these arguments (although we have changed the order in which they occur in Laudan's article.) What is distinctive about these arguments is that none of them depends on controversial theses about observation and meaning. Rather, their focus is on the criteria used to choose between competing theories and paradigms. In short, all these arguments concern epistemic values.

PROBLEM WEIGHTING

Part of the innovation wrought by Kuhn lies in his insistence that we should assess theories, not by looking at their empirical or observational consequences, but rather by seeing how good they are at solving problems. Designing a theory to agree with observation is, Kuhn thinks, a relatively simple task, especially if one is not fussy about what the resulting "theory" looks like. But, he insists, the point is not first and foremost to devise theories that agree with observation but to devise theories that answer questions and solve problems; the unit of scientific achievement is not the true observational consequence but the successful puzzle solution. Good theories belong to paradigms that lead to a progressive accumulation of solved problems. Often the key issue in the ferment preceding a scientific revolution is whether a rival paradigm is a better problem solver or has the promise of being so once the full energies of science are devoted to its development. But then, Kuhn points out, we run into a difficulty. For a characteristic feature of many scientific revolutions is that the new paradigm cannot solve all the problems of its predecessor. Some problems that were solved by the old paradigm are left unsolved or simply ignored by its successor. So, Kuhn argues, no algorithm or calculation can determine a rational choice between paradigms. The scientist's choice depends on the

relative weight assigned to particular problems, and this weighting will vary from one paradigm to another.

SHIFTING STANDARDS

According to Kuhn, paradigms include standards for assessing theories, and these standards vary from paradigm to paradigm. For example, one group of scientists might insist that no theory should be accepted unless it makes novel predictions, whereas another group might extol a theory's ability to give a unified explanation of diverse phenomena, even though all of those phenomena were previously known. The reception of Darwin's theory in the nineteenth century depended, in part, on just such a disagreement about methodology. Judgments about the relative merits of theories vary depending on the set of standards used. No higher court can adjudicate these disagreements. So, here again, Kuhn concludes that there is no single, rational way of resolving paradigm conflicts.

THE AMBIGUITY OF SHARED STANDARDS

As Kuhn has emphasized in his later writings (see "Kuhn's Second Thoughts . . ." below), some epistemic values, some views about theoretical virtue, are widely shared by scientists. But, he thinks, appeal to these shared standards cannot suffice to determine the outcome of paradigm debates. One important reason for this is the differing interpretations given to these standards. For example, proponents of rival paradigms might agree that explanatory power is desirable but disagree about what constitutes an explanation. Just such a case arose in the early eighteenth century in connection with Newton's gravitational theory. Newton's theory gave a unified lawlike treatment of the phenomena with plenty of predictive power. But Newton's European critics (especially Leibniz and the followers of Descartes) did not regard Newton's theory as having explained anything. In their view, explanation required the provision of a plausible (or at least possible) mechanism or process, that obeyed the laws of mechanics and could causally bring about the phenomena to be explained. Newton and his followers agreed that, by relying on instantaneous action at a distance,⁷ Newton's theory probably could not be reconciled with any such mechanism. So the parties agreed that good theories should explain, but they differed markedly in their understanding of what explanations are.

Another example of the ambiguity of standards is provided by the Copernican revolution. (Kuhn's first book was on the Copernican revolution, and Kuhn frequently appeals to this episode as a source of illustration and support for his views. The accuracy and legitimacy of these appeals is examined in the McMullin article, discussed below.) Both sides in the debate regarded simplicity as an important and desirable property of a scientific theory, but they disagreed about how simplicity should be

measured. Ptolemaic astronomers complained that the Copernican theory was no simpler than the Ptolemaic theory, because it used just as many (and, in some cases, more) circles to model the planetary orbits. Copernicans (such as Kepler and Galileo) saw a deeper kind of simplicity in the Copernican theory that had nothing to do with circle counting but concerned instead the ability of the Copernican theory to reflect mathematical patterns in the motions of the planets much more simply than had the Ptolemaic theory. As far as Kuhn is concerned, there is no objectively right answer as to which conception of simplicity is correct. Ultimately, he thinks, these kinds of preferences are matters of taste. Since differing interpretations of epistemic values are, Kuhn thinks, inevitable and irresolvable by rational debate, he concludes that shared standards cannot provide a common ground for settling paradigm conflicts.

THE COLLECTIVE INCONSISTENCY OF RULES

Typically, a paradigm will include several different methodological rules and standards for judging theories. These might include a preference for theories that are quantitatively accurate in their predictions, of wide explanatory scope, and parsimonious in their postulation of unobservable entities and processes. There is no guarantee that even the best available theory will satisfy all these requirements: typically, some requirements are satisfied to a greater degree than others, and some may not be satisfied at all. Now suppose that a pair of rival theories T_1 and T_2 are being judged, either by a single paradigm or by rival paradigms sharing the same rules and standards. Assume that the first theory, T_1 , gives more accurate predictions than the second theory, T_2 , but is inferior with regard to the scope of its explanations and has roughly the same degree of simplicity. The first rule—prefer theories that make quantitatively accurate predictions—favors T_1 ; the second rule—prefer theories of wide explanatory scope—favors its rival, T_2 . The entire set of rules thus gives conflicting advice, and Kuhn thinks that there is no uniquely rational way of settling the matter. Again, he concludes that scientists must inevitably have recourse to personal, subjective, and psychological factors to tip the balance one way or the other. If two scientists reach opposite verdicts, neither can be accused of being irrational. The choice between the two theories is, in Kuhn's view, extrarational; it cannot be resolved simply by appealing to rules. So even if scientists share the same rules and even if they agree about how those rules should be interpreted, paradigm choice (and even theoretical choices within a given paradigm) will not be determined by those rules.

2.4 | Kuhn's Second Thoughts about the Rationality of Scientific Revolutions

In "Objectivity, Value Judgment, and Theory Choice," Kuhn responds to those critics who charge that *The Structure of Scientific Revolutions* depicts scientific revolutions as irrational affairs governed largely by subjective preference, idiosyncrasy, and social pressure. Kuhn denies that he said any such thing, even in the penultimate chapter of that book (chapter 11, "The Resolution of Revolutions") in which he compares scientific revolutions to gestalt switches and conversion experiences. In his defense, Kuhn makes three main points.

■ SHARED EPISTEMIC VALUES

He does not deny but in fact affirms that there are some important methodological rules—shared values—that all paradigms share. Kuhn goes so far as to claim that these values are essential to the scientific enterprise, thus qualifying them as constitutive values on his view.⁸ Insofar as paradigm debates are conducted solely within the confines of these shared values, about which there is total agreement (at least in broad outline), these debates are rational and based on reasoned argument.

■ INTERSUBJECTIVE AGREEMENT

Kuhn still denies that there can be objective progress in science. He rejects cumulativeness (which he thinks is essential to progress) and dismisses the idea that we can show that our theories are getting closer to some objective, theory-independent truth.⁹ Nonetheless, he adds, this does not mean that science lacks objectivity. For objectivity consists, not in the correspondence of our theories to the world, but in the intersubjective agreement about those theories among members of the scientific community, based on their shared values. Kuhn identifies objectivity and rationality with a special sort of social consensus, a consensus based on the values that make science what it is.

■ NONRATIONAL FACTORS

For the reasons sketched in the previous section ("Six Kuhnian Arguments for Relativism"), Kuhn continues to insist that paradigm debates, by their very nature, cannot be settled solely by an appeal to evidence, reason, and argument. Reason is not irrelevant to these debates, and considerations based on shared values often play a major role in marshalling support for a new paradigm. But shared values can only take us so far, and rival paradigms usually involve other values that are not shared. So when scientists decide to switch (or not to switch) paradigms, nonrational factors play a vital role. One of the major failings of traditional philosophies of science, in Kuhn's view, is that they have falsely assumed that scientific

reasoning about which theories to accept is entirely governed by universal, context-independent rules. Philosophers of science then compound their error by thinking that science is objective only because it is completely governed by these rules.

SHARED EPISTEMIC VALUES

In this section, we begin to discuss Kuhn's first point, that all paradigms share certain fundamental values that are essential to the nature of science. (Further discussion and criticism of all three of Kuhn's points appears in the sections on McMullin and Laudan.) Three separate questions can be raised about shared epistemic values. First, what are the values that Kuhn deems essential to science, and is Kuhn's characterization of them adequate? Second, is it true that all paradigms share these values? And, third, is it true, as Kuhn claims, that rational debate during paradigm conflicts is limited solely to arguments based on shared values, or can there be a rational debate about which values to adopt?

We shall focus on the first of these questions. In "Objectivity, Value Judgment, and Theory Choice," Kuhn lists five shared epistemic values: accuracy, consistency, scope, simplicity, and fruitfulness.

■ ACCURACY

Accuracy is a matter of the theory agreeing with experiments and observations that have already been made. Presumably, accuracy can vary in degree depending on the number of such agreements and the relative precision with which the theory predicts them. Characteristically, in his discussion of accuracy, Kuhn lumps together both prediction and explanation, and he allows for both quantitative and qualitative agreement. Kuhn's discussion makes clear that he understands the accuracy of a theory to concern not merely agreement with experience (what Kuhn calls *matching*), but also explanation (what Kuhn calls *accounting for* a phenomenon or regularity, although as just noted, Kuhn often treats explanation and prediction as if they were the same thing). In Kuhn's terminology, the key to judging the accuracy of a theory is to identify the number and importance of the problems it solves. This view acknowledges the qualitative dimension of accuracy (since not all puzzle solutions are quantitative) and the ambiguity of accuracy as an epistemic value (since opinions may differ about which are the most important problems for a theory to solve).

■ CONSISTENCY

Consistency has two distinct aspects: internal and external. A theory is internally consistent when it is free from logical contradiction. A theory is externally consistent when it does not contradict any other currently accepted theory. For example, when first introduced in the sixteenth century, the Copernican theory was inconsistent with Aristotelian physics, the then-

accepted theory of motion and its causes. Similarly, Bohr's theory of the atom, which required electrons to revolve in stable orbits around a positively charged nucleus, was inconsistent with Maxwell's electrodynamic theory, according to which all accelerating charges must radiate and lose energy.

■ SCOPE

Scope concerns the breadth of the theory's logical consequences. Under the rubric of scope, Kuhn emphasizes the special importance of a theory's having surplus content: the theory should have consequences that go far beyond the set of observations, laws, and subtheories that it was originally designed to explain. (For more on the question of whether so-called novel predictions have special significance for confirmation, see chapter 4.)

■ SIMPLICITY

Simplicity has been interpreted in several different ways. Kuhn's official definition equates simplicity with unifying power: a simple theory, he contends, brings order to phenomena that would otherwise be considered separate and isolated. But this seems to characterize simplicity as another aspect of scope and to ignore the essential core of simplicity, namely, parsimony. Two theories might both give unified treatments of a wide range of phenomena (and thus have equally wide scope), but one of them might use fewer assumptions than the other and thus be judged simpler. Something like this latter notion of simplicity was crucial for the Copernican revolution. Unfortunately, Kuhn's own comments about that revolution in this context are subtly misleading. After remarking that the Copernican theory required just as much labor to compute planetary positions as did the Ptolemaic theory (and thus, by that measure of simplicity, was no simpler), Kuhn observes that, in order to explain gross qualitative features of planetary motion such as retrograde motion and the restricted elongation of the inferior planets, the Copernican theory required only one circle per planet, whereas Ptolemy needed two. What Kuhn says is literally true, but it misses the point. Granted, Copernicus required only one circle per planet to explain the phenomena in question, but the explanation in each case crucially involved another circle, namely, the earth's orbit around the sun. And in the Copernican theory, the earth, too, is a planet. So if we count the total number of planetary circles needed to explain these phenomena, we get the same answer (two) for both theories. But still, Copernicus's theory was simpler than Ptolemy's. Why? Because from the single postulate that the earth and all the other planets move in roughly circular orbits around the sun, all the phenomena follow as immediate geometrical consequences, without the need for any further assumptions. In Ptolemy's theory, additional postulates were needed to tie the motion of each planet to the motion of the sun in a manner that, in the framework of Ptolemy's theory, was completely arbitrary. Thus, Co-

pernicious's theory was simpler than Ptolemy's because it derived the same observable phenomena from a smaller set of assumptions. (For further criticisms of Kuhn's interpretation of the Copernican revolution, see the discussion of McMullin in the next section of this commentary.)

■ FRUITFULNESS

Fruitfulness, as Kuhn remarks, has often been underemphasized in traditional accounts of science. As an epistemic value used for making decisions between rival theories, the issue is not fruitfulness per se, but the promise of fruitfulness. What counts is not how successful a theory will be in leading to the discovery of new phenomena and laws (and, more generally, in solving problems) but how promising the theory appears to be right now, before that promise is realized (or not, as the case may be). Because fruitfulness (or fertility, to use McMullin's term) concerns our hopes for the future, Kuhn (in *The Structure of Scientific Revolutions*) contends that it involves an element of what he calls *faith*, and thus to a greater extent than the other epistemic values, it eludes rational evaluation.

2.5 | McMullin's Criticisms of Kuhn

In "Rationality and Paradigm Change in Science," Eran McMullin documents the ways in which Kuhn has toned down his subjectivist rhetoric since *The Structure of Scientific Revolutions*. On the one hand, Kuhn has emphasized the extent to which paradigm debate can be rational insofar as these debates are based on shared values. Kuhn assumes that values such as accuracy, scope, and the like are fixed and permanent features of science. But, on the other hand, Kuhn (as we saw earlier) continues to insist that these values can never determine the outcome of scientific revolutions. Furthermore, Kuhn persists in denying that any objective notion of progress (such as getting closer to the truth) can be applied to science across revolutionary divides. As McMullin notes, Kuhn's denial of progress follows from his conviction that even the shared epistemic values are ultimately arbitrary and subjective. Specifically, Kuhn thinks that it is impossible to show that these values are connected in any necessary way with truth or the likelihood of truth.

McMullin has three main concerns about Kuhn's later views. First, how plausible is Kuhn's claim that the epistemic values he identifies are shared by all scientific paradigms? Second, is Kuhn correct in insisting that epistemic values cannot be given any rational justification? Third, can Kuhn ignore the question of scientific realism (crucial to his denial of objective progress) and still make sense of scientific revolutions such as the Copernican revolution? More generally, is it sufficient for Kuhn simply to accept, as a sociological fact, the epistemic values that scientists

employ in judging theories without offering any deeper rationale for those values?

SHARED VALUES

With regard to the first of these questions, McMullin notes that Kuhn counts as revolutions a remarkably wide variety of scientific changes, ranging from the shallow (e.g., the discovery of X rays) to the radical (e.g., the Copernican revolution), with many in between (e.g., the victory of Lavoisier's oxygen theory over the phlogiston theory¹⁰). If one looks only at episodes from the shallow end of this range, then claiming that these revolutions left epistemic values largely untouched is quite plausible. While some of them involved fundamental shifts in ontology and concepts, the standards by which theories are judged changed little or not at all (although, of course, advocates of rival paradigms might well disagree about the interpretation and application of those standards). But *deep revolutions* (as McMullin calls them) do involve debates about standards. The very notion of what constitutes a good theory and what qualifies as evidence was at issue in the Copernican revolution and, to a lesser extent, in the quantum revolution of the twentieth century. Indeed, McMullin thinks that disagreements about standards and the merits of alternative theories can occur even within what Kuhn would classify as normal science. (McMullin cites paleontology, high-energy physics, and planetary science as examples.) Thus, disagreements about theories and standards are not limited solely to confrontations between incommensurable paradigms. They can occur, to a greater or lesser extent, even within a prevailing paradigm. A close look at the history of science reveals that the distinction between normal science and a scientific revolution is more a difference of degree than a difference of kind.

Given that scientific revolutions can and do differ in depth, why should Kuhn think that some standards (epistemic values) are immune to change? Why should the features that Kuhn identifies as shared values be permanent features of science, shared by all scientific paradigms, while other criteria for choosing among theories can change with paradigms? McMullin discusses Kuhn's attempts to answer these questions and judges his answer unsatisfactory. Kuhn himself denies that science has a fixed essence, a set of necessary and sufficient conditions that distinguish it from other human activities. Rather, in the manner of Wittgenstein's *Philosophical Investigations*, Kuhn regards our idea of science as a cluster concept. For a theory or paradigm to be scientific does not entail that it must satisfy all the values in the cluster, only that it satisfy (to some degree) a fair number of them. And activities other than science might share one or more of these values. But, as McMullin points out, this picture (of a cluster of values associated with the activities that we now recognize as scientific) is consistent with the possibility that values could be added to

or deleted from that cluster in the future in such way that the cluster as a whole would retain its special identification with science. The fact that a particular cluster of criteria are constitutive of science at present does not imply that the membership of this cluster must be permanent and unchanging.

THE JUSTIFICATION OF VALUES

Must all paradigms satisfy the epistemic values singled out by Kuhn? There is one sort of answer to this question that Kuhn thinks is impossible to give—namely, a philosophical argument connecting these values with the aim (or aims) of science. To simplify matters, let us focus our attention on truth as one of the main aims of science. Kuhn says that we cannot give any philosophical justification for the epistemic values of science because, if we could, we would have solved the problem of induction. (The problem of induction is discussed in some detail in chapter 4.) For our present purposes suffice it to say that by “solving the problem of induction” Kuhn means “showing that our criteria for theory choice lead to true theories or to theories that have a high probability of being true.”

Like Kuhn, McMullin is not sanguine about the possibility of solving the problem of induction by *a priori* philosophical reasoning. But McMullin thinks that the problem of induction might be solved, at least to a limited extent, by an appeal to experience. As he explains in the section of his paper entitled “How Might Epistemic Values Be Validated?” (pp. 128–31), some epistemic values of science (such as simplicity and fertility) are subordinate to the primary goals of predictive accuracy and explanatory power. Simplicity and fertility are not valued for their own sake or as ends in themselves but as means to further ends. Even if we cannot prove that predictive accuracy and explanatory power are indicators of truth, it is still an important empirical discovery that theories of greater simplicity and fertility tend to be more reliable predictors and better explainers. Indeed, McMullin thinks that it was an empirical discovery of the Copernican revolution that it was possible for an astronomical theory to be both an accurate predictor *and* a good explainer. (Prior to Copernicus, astronomers in the Ptolemaic tradition had pursued predictive accuracy at the expense of explanation, arguing that this was an inevitable limitation of their discipline. Explanation was relegated to the realm of Aristotelian physics.) So even if we cannot demonstrate that theories that are good predictors are more likely to be true, and even if we cannot prove that theories that have been good predictors in the past are likely to remain good predictors in the future, we can still give a rational justification of some of our epistemic values (such as simplicity and fertility) by appealing to the connection, learned from experience, between simplicity and predictive reliability. Prior to the Copernican revolution, simplicity and fertility were seldom included among the criteria of a good scientific theory.

One of the important things that Copernicus and his followers taught us was that, as a matter of empirical fact, simplicity and fertility are reliable indicators of predictive and explanatory success. So to a limited extent, we can justify some of our epistemic values by appealing to the lessons of history and experience.

RATIONALITY AND REALISM

Kuhn is not a scientific realist. He scrupulously avoids making any inference from a scientific theory satisfying our criteria of a good theory to that theory being true or even close to the truth. Kuhn evinces a kind of metaphysical queasiness about talking in a literal and straightforward manner about theories being true or the entities that they postulate being real. As McMullin points out, Kuhn's work is written from the standpoint of an instrumentalist. Kuhn thinks that he can set aside the issues of objective truth and real, theory-independent existence. As long as a theory is good at solving puzzles, no further questions need arise, as far as Kuhn is concerned. (One of Kuhn's main arguments against scientific realism, based on the radical shifts in ontology that often occur when one theory replaces another, is presented in detail in Laudan's paper, "A Confutation of Convergent Realism," in chapter 9.)

McMullin, on the other hand, is an unabashed scientific realist. Although he is understandably reluctant to embark on a full-fledged defense of scientific realism in his paper, he does raise serious questions about Kuhn's ability to explicate the rationality of scientific revolutions without taking realism seriously. McMullin is especially critical of Kuhn's account of the Copernican revolution. In his book *The Copernican Revolution*, Kuhn compares and contrasts the Ptolemaic and Copernican theories. He concludes that both theories had about the same predictive accuracy and that both were able to explain the major phenomena of planetary motion. Kuhn admits that astronomers like Copernicus, Kepler, and Galileo were attracted to Copernican theory because its explanations were more "natural," but he regards this preference as merely a matter of taste, with no objective or epistemic significance.

McMullin shows that Kuhn is dead wrong when he claims that the Copernican and Ptolemaic theories had the same degree of explanatory power. True, the two theories had about the same predictive accuracy, but the Copernican theory could explain many things that the Ptolemaic theory could not. McMullin gives several examples of such phenomena, including restricted elongation—the fact that Mercury and Venus (the so-called inferior planets) are never seen more than a certain angular distance from the sun—and various features of retrograde motion, such as the fact that a planet undergoes retrograde motion only when that planet, the earth, and the sun lie in a straight line. These phenomena are immediate consequences of the Copernican theory's basic postulates.¹¹ By an appro-

priate adjustment of parameters, the Ptolemaic theory could represent these facts, but it could not explain them. For Ptolemy's theory, they were unexplained coincidences; for Copernicus's theory, they were necessary consequences of the basic structure of our planetary system.

Copernicus, Kepler, and Galileo took the explanatory success of the Copernican theory to be a powerful argument for its truth. They reasoned that it could not be an accident that their theory was able to explain so much; only by uncovering the true causes of the observed motions of the planets could the Copernican theory account for those motions in such an elegant, unified fashion. As McMullin emphasizes, merely "saving the phenomena" was not the issue. Any number of theories could be devised, at least in principle, that could do as good a job as the theories of Ptolemy and Copernicus in predicting where the planets would be seen on any given night. Unlike predictive accuracy, explanatory power was an indicator of truth. Kuhn's instrumentalism makes him unwilling to recognize the important distinction between prediction and explanation; as a consequence, he misportrays the rational basis on which Kepler and Galileo preferred the Copernican theory. The Copernicans preferred their theory not primarily because it was a better predictor or a better problem solver than its rival, but because they believed it to be true. And they believed it to be true largely because of its tremendous explanatory power.

McMullin also argues for the epistemic significance of other theoretical virtues, such as fertility. Again, his point is that these virtues make sense only when seen as indicators of truth. Only when we talk about truth and embrace scientific realism, can we appreciate why scientists take these virtues so seriously and why they have played such an important role in the history of science. For example, one of the most persuasive arguments for the Copernican theory was the runaway success of the program of research to which it led. In part, this success stemmed from the fact that the Copernican theory (unlike its Ptolemaic rival) enabled astronomers to calculate the distance of each planet from the sun (as a function of the distance between earth and the sun). This information about distances was crucial to the advances made by Kepler and, eventually, Newton. Again, the Copernican theory's ability to generate new and important lines of research was valued not merely for its own sake or on pragmatic grounds (by promising employment and possible fame to eager young scientists) but primarily on epistemic grounds, as an indicator of truth. Only a theory (or, more generally, a paradigm) that is true or close to the truth could possibly be successful in producing such a cornucopia of wonderful new results. (Whether such arguments from the success of theories to their probable truth are plausible is discussed in detail in chapter 9, "Empiricism and Scientific Realism.")

2.6 | Laudan's Criticisms of Kuhn

Laudan's criticisms of Kuhn in his "Dissecting the Holist Picture of Scientific Change," differ markedly from those of McMullin in part because Laudan does not share McMullin's enthusiasm for scientific realism. Consequently, he evaluates Kuhn's claims at the level of problem solving without raising the issue of realism. But Laudan agrees with McMullin that Kuhn has drawn false conclusions about the nature of scientific rationality using flawed arguments. Before discussing Laudan's specific criticisms of Kuhn, it will be helpful to contrast two accounts of scientific rationality: the hierarchical model and the reticulation model. (The terms are Laudan's, and his account of them can be found in chapters 2 and 3 of his book *Science and Values*; "Dissecting the Holist Picture of Scientific Change" is chapter 4 of that book.) Part of Kuhn's error, Laudan argues, is that he has uncritically adopted the hierarchical model of scientific rationality, which is fundamentally flawed.

THE HIERARCHICAL MODEL OF SCIENTIFIC RATIONALITY

Paradigms can be thought of as having three sorts of components, each at a different level. Laudan calls these levels the factual, the methodological, and the axiological. At the factual level, paradigms provide a conceptual framework, a general account of the entities and processes that supposedly populate the world. Also at this level are particular scientific theories, framed in terms of this ontology. The next level up is the methodological, which contains rules for choosing among theories. These methodological rules can include specific directives about what kinds of techniques and instruments to use in designing experiments, rules about what constitutes confirming evidence for a theory, and general statements of epistemic values such as testability and simplicity. Although it is convenient to talk simply of rules in this context, these must be understood to include statements of general preference such as, "Other things being equal, prefer theories that are simpler than their rivals." As Kuhn insists, and Laudan agrees, rules of the latter type function as what Kuhn calls values: rather than dictating a specific choice among theories, they merely specify the virtues of a good theory, leaving the individual scientist to judge how they should be balanced against other, similar, values. At the third and highest level, the axiological, are the aims and goals of science.

This model is hierarchical in the following sense: when there is disagreement at one level, the disagreement is rationally resolved (if at all) by going to the next highest level. Thus, to choose among competing theories at the factual level, appeal is made to rules at the methodological level. It is the methodological rules that provide the rationale for decisions about theories. But what if disagreement persists because of a difference

of opinion about which rules to adopt? In that case, scientists must have recourse to the axiological level. For disputes about methodological rules can be rationally resolved only by specifying the goals that the rules are supposed to serve. If scientists all shared the same views about the aims of science, then, in principle, the dispute could be settled by seeing which rules would best promote those aims. But if scientists disagree about aims, then no rational resolution is possible. If, for example, one group of scientists insists that empirical adequacy is the most important goal of scientific theorizing, while another group maintains, with equal vigor, that science should aim at providing intelligible explanations, then the two groups have no objective, rational way of settling their dispute. According to the hierarchical model, the rational resolution of axiological disagreements is impossible because it would require going to a higher level, and there is no such level.

Admittedly, this is a highly idealized and simplified sketch of the hierarchical model, but it suffices to illustrate two of its central features. First, the direction of justification is strictly top-down and linear: always from a higher level to a lower one, never the reverse. Second, at the highest level, the axiological, rational justification ceases. On the hierarchical model, views about the goals and aims of science are necessarily arbitrary and subjective. Although they are viewed as the ultimate source of justification for the decisions made at the lower levels, they themselves cannot be justified.

THE RETICULATIONAL MODEL OF SCIENTIFIC RATIONALITY

According to Laudan, a misleading picture of scientific rationality pushes Kuhn to conclude that reason and argument play only a limited role in scientific revolutions. For, once one accepts the hierarchical model, disagreements about the criteria for assessing theories will be incapable of rational resolution if, as Kuhn supposes, rival paradigms have competing notions of the aims and goals of science. Another error contributing to Kuhn's embrace of relativism is his holism, his insistence that paradigms are seamless wholes. Kuhn assumes that paradigms are package deals that can be accepted or rejected only in their entirety. Specifically, he denies that scientists can evaluate or choose among the individual components of a paradigm, especially at the methodological and axiological levels.

Laudan advocates the reticulational model of scientific rationality. Its two main features—its antiholism and its nonlinear conception of justification—directly contradict the hierarchical model. According to the reticulational model, the components of a paradigm can be discussed, rejected, or accepted piecemeal. Like McMullin, Laudan criticizes Kuhn's assumption that paradigms have a kind of holistic unity that would make such a dissection impossible. He supports his case with historical examples in which scientists clearly did pick and choose elements from the para-

digms they were debating. Contrary to Kuhn, changes at one level do not have to be accompanied by changes at a higher level: for example, methodological rules can change in response to a discovery that the old rules are not the best way to realize the aims of science; it is precisely because scientists have not changed their view of the aims of science that they modify their rules.

Laudan's notion of justification is far richer than the strictly top-down, linear notion of the hierarchical model. On his model the flow of justification can be bottom-up as well as top-down, from the factual to the methodological as well as the other way around. He offers historical examples of how the success (and, in some cases, failure) of scientific theories led scientists to adopt (or abandon) particular methodological rules. Of particular importance is Laudan's claim that there can be, and occasionally has been, rational debate concerning the appropriate aims and goals of science. Laudan gives as an example the abandonment of the Aristotelian ideal that science should aim at certain, infallible knowledge. By the end of the nineteenth century just about every scientist agreed that science could not possibly achieve that goal but should settle instead for highly probable, but fallible, belief. As Laudan points out, this major change at the axiological level was associated not with any one particular paradigm, nor with any one particular area of science. Rather, the shift grew out of the awareness that many of the best scientific theories (whether in physics, biology, or chemistry) involved hypothetical elements that were impossible to prove with certainty. In short, the widespread success of a whole range of theories at the factual level led to the realization that the traditional Aristotelian ideal could not be attained. Since the goal of certainty could not be achieved, it was rational to abandon it.

THE FLAWS IN KUHN'S ARGUMENTS

In general, a philosophical argument can fail for three reasons: either it is invalid (i.e., its conclusion does not follow from its premises); it has at least one false premise (in which case the argument, even if valid, is unsound); or it is circular (in which case it begs the question by taking for granted the very point that it purports to prove.) In the second half of his paper, Laudan undertakes to expose the flaws in Kuhn's arguments for his conclusion that methodological rules and standards can never determine the choice between paradigms. Laudan refers to Kuhn's conclusion as the thesis of *local underdetermination*. The alleged underdetermination is local because it concerns the choices made by scientists between the few paradigms that have actually been put forward for serious consideration. *Global underdetermination* (discussed at length in chapter 3) is the far more sweeping thesis that our rules can never pick out a particular theory from among all the possible alternatives as the only one that is supported by the available evidence. Suffice it to say for now that the two

theses are different and that even if (as Laudan argues) Kuhn is wrong about the local thesis, it might still be the case that the global thesis is true. (Hence the need to address the global thesis and related issues in the next chapter.)

Kuhn's six arguments—appealing to theory-ladenness of observation, meaning variance, problem weighting, shifting standards, the ambiguity of shared standards, and the collective inconsistency of rules—were summarized earlier (pp. 219–26). The first two of Kuhn's arguments were criticized above. Laudan's response to the last three of these arguments is to accuse them of being unsound. Each of them, he says, has at least one false premise: standards do not always change from one paradigm to another; some shared standards (such as consistency and novel prediction) are not ambiguous; and not all sets of methodological rules give conflicting advice.

Laudan's criticism of the problem-weighting argument is more complex. The nub of Kuhn's argument is that proponents of rival paradigms must disagree about which problems are the most important even when they agree about the rules and standards by which theories should be assessed. As Laudan explains, *importance* can mean two different things in this context. A problem might be judged important for social or economic reasons; or simply because a particular scientist is interested in solving it. That would be a nonepistemic sense of *importance*, since it would have no direct bearing on whether a theory that solves the problem is true, confirmed, or rationally credible. Contrast that with a problem that is considered important because its solution would confirm a theory. Such a problem would be epistemically important; it would have what Laudan calls *probative* significance. The probative significance of a problem is an objective matter, and claims about it will be defended or attacked by appealing to epistemic and methodological rules—rules that deal with questions of empirical support, evidence, and confirmation. So in this second sense of *importance*, disputes about which problems are the important ones for a theory to solve will not necessarily degenerate into mere differences of subjective opinion. Thus, Laudan condemns Kuhn's problem-weighting argument as invalid. Once we focus on the epistemic sense of *importance*, it simply does not follow that proponents of rival paradigms must disagree about which problems are important or, when such disagreements do occur, that such conflicts cannot be resolved by reasoned argument based on shared standards.

Laudan concludes by discussing a case from the history of chemistry that both Kuhn and Gerald Doppelt have appealed to for support of their position. Laudan admits that Dalton's atomic theory failed to address many of the questions that older chemical paradigms had tried to answer, for Dalton's interests lay elsewhere. But, Laudan counters, Dalton's assessment of which problems of chemistry are central reflects his pragmatic interest in certain problems rather than others. This subjective interest

should not be confused with the judgment about which problems have the greatest epistemic value. Indeed, as Laudan points out, the problems that the advocates of a paradigm are most eager to solve often have little or no confirming value. For according to some philosophers, a problem solution does little or nothing to confirm a theory if the solution is one that the theory was deliberately designed to yield. (For an extended discussion about the relation between confirmation and novel predictions, see chapter 4.)

The evaluation of Kuhn's work by McMullin and Laudan is largely negative. Both authors agree that Kuhn has failed to establish his more sweeping claims about the irrationality of scientific revolutions. Neither the history of science nor Kuhn's philosophical arguments show that scientific revolutions cannot be resolved by rational argument based on evidence and shared rules. By treating paradigms as indivisible wholes and by failing to appreciate the ways in which rules and aims can be rationally debated, Kuhn has seriously underestimated the role of reason in paradigm debates. In passing this verdict on Kuhn, it is important to remember that McMullin and Laudan are not claiming that reason, rules, and evidence always suffice to determine the outcome of such debates. Rather, they are objecting to Kuhn's assertion that they can never do so. Thus, according to McMullin and Laudan, Kuhn's thesis that subjective, psychological, and rhetorical factors must play a leading role in all scientific revolutions is ill supported and false.

Many philosophers of science agree that the more extreme, subjectivist elements in Kuhn's picture of science (especially the picture presented in the first edition of *The Structure of Scientific Revolutions*) are exaggerations. Nonetheless, many of those same philosophers of science find much of value in Kuhn's work. This is especially true of one aspect of Kuhn's approach that we have not yet discussed, namely, his view that science is essentially a social activity. One philosopher who has tried to develop this theme in Kuhn's work and relate it to traditional philosophical concerns about objectivity and rationality is Helen Longino.

2.7 | Longino on Contextualism and Objectivity

In "Values and Objectivity," Longino begins by distinguishing two notions of objectivity. The first concerns the outcome of scientific inquiry, namely scientific knowledge; the second focuses on the process by which that outcome is generated, namely scientific method. According to the first notion, science is objective to the extent that it is true—to the extent that it describes the world as it really is. According to the second notion, science is objective to the extent that its methods, especially its criteria for assessing theories, are neither arbitrary nor subjective. As Longino points

out, it is commonly believed that science is objective in the first sense because it is believed to be objective in the second.

THE LOGICAL POSITIVIST VERSUS THE CONTEXTUALIST ACCOUNT OF SCIENTIFIC METHOD

Longino contrasts two accounts of scientific method: the logical positivist account and the contextualist account of "historically minded wholist critics," such as Kuhn and Feyerabend. The logical positivist account divides scientific activity into two distinct phases: the context of discovery and the context of justification. The context of discovery is the period during which scientists first come up with their ideas and hypotheses. Here it is allowed that psychological and subjective factors often play a major role. But once the theory or hypothesis has been formulated and presented for assessment, then the context of justification begins. In this context, positivists claim that psychological and subjective factors play no role whatever (or, rather, that they *ought* to play no such role, since positivists are strongly wedded to the value-neutrality thesis). On one extreme version of the logical positivist view (and the only one that Longino considers), assigning degrees of confirmation to hypotheses and deciding which theories to accept (or reject) is solely a matter of applying the relevant rules to the available evidence. This rule-based assessment could be performed by a single scientist; if several scientists assess the same theories on the basis of the same evidence, they should all get the same answer. The social dimension of science is irrelevant to its production of reliable knowledge. In principle, science could be done by a single individual.

Although popular in the first half of the twentieth century, the logical positivist account of scientific method is now discredited, in part because of the criticisms of Kuhn and others. One of those criticisms is that the logical positivist picture of science does not match what has actually happened during the history of science. During scientific revolutions especially, we do not find anything like the degree of consensus about scientific theories that the logical positivist account leads us to expect. Another criticism concerns the shortcomings of the logical positivist treatment of confirmation. One of the things that Kuhn has taught us, Longino explains, is that confirmation cannot be a matter of mechanically applying rules to evidence because what an individual scientist counts as evidence for a theory in the first place will depend on the other beliefs that the scientist holds. Background beliefs and assumptions are crucial in determining which hypotheses we accept as being confirmed by which evidence. Longino calls this the *contextualist analysis of evidence*.

THE CONTEXTUALIST ANALYSIS OF EVIDENCE

In chapter 3 of her book *Science as Social Knowledge*, Longino gives the following illustration of the contextualist analysis of evidence. Suppose that an eight-year-old child has red spots on her stomach. Why should her parents take the rash as evidence for measles rather than, say, as evidence that the moon is blue? Ordinarily, the explanation would be that they think that the rash is caused by the measles virus, and there is no other likely cause of that kind of rash. Given this background belief, the rash strongly confirms their diagnosis of measles. But they could have held different background beliefs. They might, for example, have believed that a rash of this kind is caused by a gastric ailment. In that case, they would have inferred a different hypothesis from the same evidence. Furthermore, the things that they regard as evidence can be described in a variety of ways: red spots on the stomach, spots (without mentioning their color), a rash (without mentioning its location), urticaria, an itchy tummy. Which hypotheses they take to be confirmed by the evidence also depends on how that evidence is described.

Two important features of Longino's contextualist analysis of evidence need to be stressed. First, what Longino calls *evidence* is a physical object, a physical process, or, most generally, a state of affairs. This is the way that the term *evidence* is used in a court of law (where it also includes testimony). It is not the way that most philosophers of science use the term when discussing confirmation. As we shall see in chapters 4 and 5, when analyzing confirmation most philosophers of science assume that the evidence for a theory or hypothesis is a set of statements. That is, having assumed that the objects (events, processes) we observe have been described in some language, these philosophers then go on to analyze confirmation as a relation between observation statements and theoretical statements. While the distinction between the objects of observation (what Longino calls *evidence*) and observation statements (what most other philosophers of science call *evidence*) might seem pedantic, it is crucial for understanding the nature of Longino's project. For many philosophers of science would agree with Longino that the first step, the step from sensory stimulation to observation report, depends on the observer's background beliefs and training, and the language that person decides to use. For example, the words uttered when looking through a telescope trained on the moon will vary from individual to individual. To that extent, confirmation involves a subjective component. But acknowledging this component is consistent with the view that the confirmation relation between observation statements and theoretical statements is entirely objective. Longino stresses the subjective element in the first step because she is interested in the entire causal and social process by which scientists make judgments about theories. As she describes it, her project is to *explain* this kind of scientific activity by understanding the mechanisms at work. Like

Kuhn, Longino is more concerned with explaining scientific behavior (and exploring its consequences for objectivity and rationality) than with giving formal analyses of isolated concepts.

Second, in presenting her contextualist analysis of evidence, Longino is scrupulous about respecting the important distinction between two questions: "Does this piece of evidence confirm this hypothesis?" and "Why do scientists take this piece of evidence to confirm this hypothesis?" Longino confines the contextualist analysis of evidence to answering the second question, deliberately leaving the first question to one side. One reason Longino avoids the first and more traditional question is that she thinks that philosophical attempts to give formal analyses of confirmation have been failures.¹² Another reason stems from her concern with explanation. For it is unlikely that a sophisticated, formal model of confirmation—even if it should prove philosophically adequate—could play a causal role in the individual psychological processes or the collective social mechanisms by which scientists reach judgments about confirmation. Since her primary goal is to explain scientific behavior, it is the second question, not the first, that is relevant to her project.

Nonetheless, it is still possible that a convincing philosophical answer to the first question might be found. For example, one of the attractions of the Bayesian approach to confirmation (explored in chapter 5) is that it provides a formal framework within which background beliefs can play an important role. Some philosophers of science (such as Wesley Salmon) have argued that this approach provides a perfect vehicle for incorporating the subjective factors Kuhn and Longino emphasize. So even though we might agree with Longino that background beliefs play a central role in confirmation, confirmation may still be an objective relation between statements that include those background beliefs as an element.

In summary, according to the contextualist analysis of evidence, confirmation (or, more precisely, what we take to be confirming evidence for a theory or hypothesis) depends on background beliefs and assumptions, what Longino calls *contextual values*, which can vary from individual to individual. How we describe the evidence and which hypotheses we judge to be confirmed by that evidence (and to what degree) all depend on contextual values.

CRITERIA FOR OBJECTIVITY

Despite her emphasis on contextual values in science, Longino is far from denying that science is objective. In fact, she adamantly rejects the charges of rampant subjectivism and irrationality that have been hurled against Kuhn. She thinks that critics of Kuhn have made two fundamental mistakes: they have failed to distinguish between the two senses of objectivity—process versus product—discussed at the beginning of her article, and they have failed to take seriously the social character of scientific

activity. Longino regards the latter as the key to guaranteeing some degree of objectivity in science.

As Longino points out, the contextualist analysis of evidence raises a serious question: if all judgments about confirmation and evidence depend upon the subjective beliefs of the individual researcher, how can science be objective? The problem is serious because subjective background beliefs (contextual values) often escape the notice of the person who holds them. And unlike advocates of the value-neutrality thesis, Longino does not think it possible to eliminate contextual values from science.

Longino believes that the answer to the question, "How can science be objective, when contextual values play an essential role in the judgments made by individual scientists?" lies in the realization that science is a social activity organized to permit and encourage what she calls *transformative criticism*—that is, criticism with the power to change contextual values should they prove ill-founded. Thus, science can be objective only to the extent that transformative criticism is allowed to flourish. In fact, she regards this social feature of science as necessary for objectivity. Longino writes:

From a logical point of view, if scientific knowledge were to be understood as the simple sum of finished products of individual activity, then not only would there be no way to block or mitigate the influence of subjective preference but scientific knowledge itself would be a potpourri of merrily inconsistent theories. Only if the products of inquiry are understood to be formed by the kind of critical discussion that is possible among a plurality of individuals about a commonly accessible phenomenon, can we see how they count as knowledge rather than opinion. (179)

I have argued both that criticism from alternative points of view is required for objectivity and that the subjection of hypotheses and evidential reasoning to critical scrutiny is what limits the intrusion of individual subjective preference into scientific knowledge. (181)

In the remainder of her paper, Longino describes four criteria that, she argues, are necessary for transformative criticism to flourish (181):

- 1 "There must be recognized avenues [such as journals and conferences] for the criticism of evidence, of methods, and of assumptions and reasoning." Scientists should get appropriate credit for their critical activities, just as they do for their original research.
- 2 "There must exist shared standards that critics can invoke." Longino's list is similar to Kuhn's, although, like Kuhn, she denies that these standards can serve to determine choices among theories. Nonetheless, these constitutive values (or shared standards) play an

important role in constraining and correcting the influence of contextual values.

- 3 "The community as a whole must be responsive to such criticism." Longino does not think that individual scientists should recant the moment they are criticized, for science is often best served when scientists defend their work from criticism. But criticism must be capable, in the long run, of changing the beliefs of the scientific community as a whole.
- 4 "Intellectual authority must be shared equally among qualified practitioners." Longino alludes here to the ideas of the German philosopher Jürgen Habermas, who has stressed the importance of public debate and rational consensus for preventing the domination of society by one group of interests.¹³ An egregious example of the violation of this criterion was the suppression of Mendelian genetics in the Soviet Union in the 1930s.¹⁴ Longino thinks that the exclusion of women and racial minorities from science in the United States also violates this criterion.

Longino believes that these criteria are an ideal to which science should strive rather than a description of how actual science is conducted. Thus, Longino is much clearer than Kuhn in distinguishing between the descriptive and the normative in her account of science. Kuhn asserts that normal science is relatively dogmatic, reaching consensus by training scientists in one particular paradigm and discouraging dissent. Longino advocates that scientists should encourage criticism and reach consensus by rational negotiation; when science falls short of this ideal, its objectivity is compromised. Thus, Longino has a much clearer basis from which to criticize scientific behavior than Kuhn does. And Longino does precisely that in the later chapters of her book, where she criticizes contemporary biological research into sex differences and the explanation of human and animal behavior. (This research and criticisms of it are reviewed in Kathleen Okruhlik's paper, "Gender and the Biological Sciences," discussed below.)

Another feature of Longino's criteria is that she regards them as necessary conditions for scientific objectivity. But are they necessary conditions in the strict logical sense? Arguably, it is an empirical matter whether or not (or to what extent) scientific objectivity is compromised when, say, women or racial minorities are excluded from science.¹⁵ Were all branches of science, from mathematics to chemistry, less than fully objective throughout the entirety of their history because, until recently, women and minorities were almost completely excluded from them? There is also the question of whether Longino's criteria for scientific objectivity are sufficient. In her paper, "Gender and the Biological Sciences" (discussed below), Kathleen Okruhlik argues that no set of criteria concerned solely

with choosing among already formulated theories can guarantee objectivity.

A final point concerns the notion of objectivity itself. Longino's criteria are designed to promote objectivity by trying to eliminate the bias, arbitrariness, and subjectivism of individual scientists from the judgments made by the scientific community as a whole. But what is the connection, if any, between this notion of objectivity (roughly, freedom from bias) and the notion of objectivity mentioned at the outset of Longino's paper, namely, objectivity as truth? Habermas, whom Longino discusses at length in her book (see especially pp. 200–201), connects the two by *defining* truth as whatever all participants would agree about, provided they all had an equal chance to engage in free and uncoerced communication and all had equal power to make their views rationally persuasive. But, as Longino remarks, this attempt to define truth as rational consensus faces a number of powerful objections. What happens when consensus changes over time? Does this mean that truth also changes? And what about propositions concerning which there will never be consensus, not even in the long run? Are these propositions neither true nor false? Perhaps it is more plausible to regard scientific consensus not as a definition of truth, but as a reliable indicator of truth. In that case, however, we would seem to need some independent way of ascertaining the truth in order to justify our belief that scientific consensus leads to the truth. And what could that independent way possibly be when science is our best and only guide? These are difficult questions, but they must be confronted by anyone who believes that science is not merely a game played for its own sake but a reliable guide to the truth. (This issue is discussed at length in chapter 9, "Empiricism and Scientific Realism.")

2.8 | Okruhlik on the Feminist Critique of Science

In the first section of her article, "Gender and the Biological Sciences," Kathleen Okruhlik discusses several cases from modern biology in which contextual beliefs about gender (sexist and androcentric bias) have influenced scientific judgment. For example, the general belief that females are passive and males active may have prevented the recognition of evidence that the ovum plays an active role in fertilization. As Longino has stressed, the issue is not whether a particular observation was made—microvilli extending from the ovum were seen and photographed as early as 1895—but how the observation is described and, as a consequence, what evidential significance it is deemed to have. For example, it makes a difference whether what is seen under the microscope is described merely as "projections around the ovum *accompanying* the penetration of a sperm cell," or as the ovum "*clasping* the sperm and *drawing* it in."

Okruhlik summarizes the influence of contextual values by alluding to the theory-ladenness of observation, but it is probably best regarded as an illustration of Longino's contextualist analysis of evidence, discussed in the previous section of this commentary.

Other examples indicate that androcentric bias played (and presumably continues to play) a major role in evolutionary biology and anthropology by shaping the questions asked and the theories devised to answer them. Okruhlik frames this part of her discussion by referring to the thesis of the underdetermination of theory by data. She claims that, even after the data have been described in some particular way, several (perhaps, indefinitely many) empirically adequate hypotheses might still be available. If all the hypotheses are empirically adequate, then contextual values, not cognitive ones, are determining which hypothesis gets selected. Although often appealed to in discussions such as these, the underdetermination thesis is highly controversial. (This thesis is examined at length in chapter 3, "The Duhem-Quine Thesis and Underdetermination.") But despite doubts about the truth of the underdetermination thesis as a generalization about scientific theories, in some cases, such as the ones described by Okruhlik, contextual values of a sexist kind have undeniably influenced theory choice.

Okruhlik points out that contextual values also shape scientific judgment when scientists steadfastly refuse to give up a hypothesis despite the accumulation of contrary evidence. An entrenched belief that women are biologically determined to be intellectually inferior to men has prevented some (male) scientists researching sex differences from recognizing when pet hypotheses (such as the hypothesis that spatial ability is X-linked) have been refuted by the evidence. Such cases illustrate the ambiguity of falsification, or what is often referred to as the Duhem-Quine thesis (see chapter 3 for a fuller discussion).

VARIETIES OF FEMINIST CRITIQUE

Assuming that Okruhlik and the authors she cites have made a plausible case for the pervasive influence of sexist beliefs on decision making in several branches of the life sciences, what are we to make of this fact? And what implications does it have for the larger question of scientific objectivity? For example, do the cases of androcentric bias show that science is, by its very nature, sexist? Or can these cases be dismissed as regrettable instances of "bad science," mere deviations from the objectivity that normally prevails in science?

At this point Okruhlik introduces a helpful trio of categories that Sandra Harding has proposed for classifying feminist criticisms of science: feminist empiricism, standpoint epistemologies, and feminist postmodernism.

Feminist empiricism regards science as essentially objective by virtue

of its methods: rules of evidence, confirmation, and falsification. If these rules were always followed correctly, "bad science" would never occur. The whole point of scientific methodology is to make the gender, race, and personality of the individual scientist irrelevant. Thus, feminist empiricists espouse the value-neutrality thesis and deny that contextual values should play any role in science.

If the arguments of Longino, Okruhlik, and others are sound, then contextual values cannot be eliminated from science and the value-neutrality thesis embraced by feminist empiricists is not plausible. But this leaves open other versions of feminist criticism in which the objectivity of science is construed, along the lines proposed by Longino, not as a feature of methodological rules, but as the outcome of a complex social process designed to encourage criticism. Okruhlik advocates this alternative in the second half of her paper.

Standpoint epistemologists such as Sandra Harding insist that contextual values are essential to science and that some contextual values are better than others. If science were done by women, from the standpoint of women, the theories produced would give us a more objective picture of the world. (Here, of course, *more objective means truer*.) As far as getting to the truth about the world is concerned, women are supposed to have certain advantages over men, either because women as a group have fewer contextual beliefs to distort their ways of seeing and thinking about the world or because the contextual beliefs of women as a group are more in tune with reality.

There are a number of problems with standpoint epistemology. First, it is difficult to specify, even in broad outline, what contextual beliefs and values women might have that would give them a natural advantage over men in discovering important scientific truths about the world. Helen Longino makes a stab at this in her paper "In Search of Feminist Epistemology" in which she assembles a list of theoretical virtues that are often emphasized in feminist writings.¹⁶ Longino contrasts this list with the set of shared cognitive values laid out in Kuhn's "Objectivity, Value Judgment, and Theory Choice." Kuhn's list is: accuracy, consistency, scope, simplicity, and fruitfulness. Longino's "feminist" list is: empirical adequacy, novelty, ontological heterogeneity, complexity of relationship, applicability to current human needs, and diffusion of power. The first item on both lists is essentially the same thing under different names. Although the accuracy (empirical adequacy) of theories is clearly relevant to their truth, it obviously cannot explain why feminist science would be epistemically superior to science as we now know it. The last two virtues on the feminist list might be relevant to the claim that feminist science would be better in some nonepistemic, moral sense (for example, by doing more to improve the lives of human beings), but it is hard to see what they have to do with truth. That leaves novelty, ontological heterogeneity, and complexity of relationship. Despite the relish with which Sandra Harding has

advocated breaking with all the assumptions on which science is currently based, it is difficult to recommend mere novelty as a guide to the truth. Ontological heterogeneity and complexity of relationship have been cited as important themes in the work of Barbara McClintock on the genetics of maize.¹⁷ But like simplicity, these values are both hard to define and difficult to connect with truth in any straightforward way.

A second difficulty with standpoint epistemology has been identified by Okruhlik and others: it is doubtful that any single standpoint is shared by all women, regardless of their age, race, sexual orientation, or economic status. Given this multiplicity of standpoints, there is a tendency for standpoint epistemology to degenerate into what Harding calls *feminist postmodernism*.

Feminist postmodernism is a variety of epistemic anarchism. All pretense to objectivity is abandoned in favor of the position that there are many different standpoints, each telling a different story about how the world is, and no one of these stories is better than any other. Science is just one possible story. For anyone who wants to criticize science on the grounds that it displays gender bias, feminist postmodernism is not an option, since it eschews all normative judgments.

OKRUHLIK'S PROPOSAL

In the second half of her paper, Okruhlik argues that the contextual values influencing the generation of theories in the context of discovery will inevitably affect the content of science in the context of justification. For the context of justification involves choosing among the available theories, and so if all those available theories are infected with gender bias, then that bias will continue to contaminate the theory we select as the best. Contrary to popular opinion, the scientific method (understood as rules, criteria, or values for theory selection) cannot, by itself, remove sexist bias from the content of science. Even if every scientific decision is perfectly rational (in the sense that it scrupulously follows all the right rules), the resulting product may still be defective because social factors have influenced the context of discovery. Thus, Okruhlik concludes, philosophers of science must acknowledge that the context of discovery has an epistemic significance and that philosophy of science should not focus exclusively on the context of justification.

In the light of her analysis, Okruhlik proposes that we should improve the objectivity of science by changing its social organization. Among other things, this means encouraging women to pursue scientific careers, eliminating sex discrimination in promotion and hiring, making sure that research by women gets the recognition it deserves, and ensuring that the voices of women scientists are heard and taken seriously when they criticize the work of their male colleagues. For, like Longino, Okruhlik locates the objectivity of science, not in its methodological rules or its cognitive

values, nor in the vain hope that contextual values could be eliminated at the level of the individual scientist, but in the community of scientists striving to encourage criticism by including a diversity of viewpoints.

As Okruhlik notes, her proposal falls somewhere between the first two types of feminist critique on Harding's list. With feminist empiricism it shares the conviction that science should aim at improving its objectivity and that established scientific method (based on the cognitive values of science) plays an important role in achieving that goal. But, unlike feminist empiricism, Okruhlik's proposal insists that contextual values influence the content of science and that the social arrangements of science in controlling that influence are of vital importance. As Okruhlik puts it, both feminist empiricism and standpoint epistemology err when they assume that the rationality of the scientific community is just individual rationality writ large or that science can be reformed at the level of individual psychology. Unlike standpoint epistemology, Okruhlik does not assume that women have distinctively feminist values that are superior to those of men. Nor does Okruhlik claim that including more women in science in the ways proposed is either necessary or sufficient for an improvement in scientific objectivity. But she thinks that such an improvement is likely, given the way science operates. In short, science will quite probably become more objective if her proposal is adopted, not by being transformed into some new kind of feminist science approached from a distinctively feminist standpoint, but by including the feminist standpoint along with others.

2.9 | Summary

In one way or another, all the readings in this chapter focus on Kuhn's *The Structure of Scientific Revolutions*. For it was Kuhn's book that made such a dramatic case for the role contextual values play in shaping scientific judgments about theories. Kuhn's notion of a paradigm and his views about the limits of cognitive values in determining the outcome of scientific revolutions have been influential and remain controversial.

As we have seen, Kuhn has two groups of arguments for his conclusion about the lack of rationality and objective progress in scientific revolutions. The first group includes the arguments from the theory-ladenness of observation and from meaning variance. Both have been widely criticized, and in his more recent publications Kuhn has toned down considerably his reliance on these arguments. The second group includes a number of arguments claiming that cognitive values alone cannot determine a scientist's decision to embrace or reject a new paradigm. Many critics have accused *The Structure of Scientific Revolutions* of portraying science as irrational. In later papers, such as "Objectivity, Value Judgment,

and Theory Choice," Kuhn has responded to these critics by emphasizing the importance of shared, cognitive values in providing a basis for reasoned argument during paradigm debates. Nevertheless, he still insists that shared values can take us only so far in deciding between rival paradigms. Ultimately, all such decisions must rest on personal, subjective, and cultural factors that are beyond the reach of rationality.

McMullin and Laudan both criticize Kuhn's second group of arguments. McMullin is especially critical of Kuhn's assumptions that all paradigms share a common set of epistemic values, that epistemic values cannot be given any rational justification, and that we can make sense of scientific revolutions (such as the Copernican revolution) without addressing the question of scientific realism.

Laudan condemns all of Kuhn's arguments as unsound and traces their defects to Kuhn's adoption of a flawed, hierarchical model of scientific rationality. In opposition to the hierarchical model of scientific rationality, Laudan advocates his own, reticulational model. One important feature of the reticulational model is that it allows for the rational justification of the components of paradigms at a variety of levels, recognizing that scientists can and do discuss, accept, and reject the components of paradigms piecemeal. Moreover, Laudan thinks that not all justification is top-down. Thus, unlike Kuhn, Laudan believes that scientists can and do engage in rational debate concerning the aims and goals of science.

Longino offers a new approach to understanding the nature of scientific objectivity by defending the contextualist analysis of evidence. Which hypotheses we take to be confirmed by which evidence crucially depends on our background beliefs and assumptions. Nonetheless, science can be objective if its social organization permits and encourages criticism of these beliefs. Thus, she concludes, scientific objectivity is a matter of degree, depending not on the methodological rules by which individual scientists judge and select hypotheses, but on the way that the scientific community is organized.

Okruhlik discusses several recent cases in the life sciences that have led feminists to accuse science of sexist bias. Okruhlik thinks that the feminists are right: in some areas of biology and anthropology, androcentric bias has compromised scientific objectivity. Of particular importance is the role that contextual values play in shaping the formation of hypotheses in the context of discovery. In this way, regardless of the rationality of our procedure for selecting among theories, contextual values can infect the content of science. Like Longino, Okruhlik thinks that the solution lies in the reform of the social organization of science, making sure that it includes feminist and other standpoints. Okruhlik compares and contrasts her proposal with three varieties of feminist critique of science—feminist empiricism, standpoint epistemologies, and feminist post-

modernism—and argues that her position falls somewhere between the first two.

■ | Notes

1. Also important is the question of how scientific theories have affected our views about morality. Of particular interest here is the effect that biological theories (such as Darwin's) have had on our ideas about egoism, altruism, and cooperation, and whether there is any sort of evolutionary ethics worthy of the name. Again, fascinating though this is, it falls outside the scope of this chapter.
2. Kuhn retracted this claim in his *Postscript*, where he concedes that "the members of all scientific communities, including the schools of the 'pre-paradigm' period, share the sorts of elements which I have collectively labelled 'a paradigm.' What changes with the transition to maturity is not the presence of a paradigm but rather its nature. Only after the change is normal puzzle-solving research possible" (*The Structure of Scientific Revolutions*, 179).
3. For criticisms of Kuhn's appeal to the history of science, see Janet A. Kourany, "The Nonhistorical Basis of Kuhn's Theory of Science," *Nature and System* 1 (1979): 46–59, and Martin Curd "Kuhn, Scientific Revolutions and the Copernican Revolution," *Nature and System* 9 (1984): 1–14.
4. The example is adapted from Carl Kordig, "The Theory-Ladenness of Observation," *Review of Metaphysics* 24 (1971): 448–84.
5. This point is made with devastating clarity in Israel Scheffler, *Science and Subjectivity* (Indianapolis, Ind.: Bobbs-Merrill, 1967), 36–40.
6. And, of course, Kuhn himself agrees with us, since, using his terminology, the accumulation of anomalies, of observations that contradict our theoretical expectations, are a prerequisite to a scientific revolution. Thus, it cannot be Kuhn's view that perception is so theory laden as to preclude such observations.
7. Newton's theory requires that the force of gravity acts instantaneously between every pair of particles in the universe, regardless of their distance apart and the emptiness of the space that separates them. Among other things, this implies that, if the sun were suddenly to disappear, the effect on the earth would be immediate and the earth would start moving in a straight line at the very instant that the sun vanishes. In modern field theories of gravity (such as the general theory of relativity) it would take about eight minutes for the effect to reach the earth (since gravitational influence travels at the speed of light). It was precisely because action at a distance seems so spooky, so "miraculous," lacking as it does any intelligible mechanism, that Leibniz and others declared it to be scientifically illegitimate.
8. "Throughout this paper I have implicitly assumed that, whatever their initial source, the criteria or values deployed in theory choice are fixed once and for all, unaffected by their participation in transitions from one theory to another. Roughly speaking, but only very roughly, I take that to be the case. If the list of relevant values is kept short (I have mentioned five, not all independent) and if their

specification is left vague, then such values as accuracy, scope, and fruitfulness are permanent attributes of science" (Kuhn, "Objectivity, Value Judgment, and Theory Choice," *Essential Tension*, 335 [114]).

9. As Peter Lipton has put it, Kuhn is "Kant on wheels." Like Kant, Kuhn distinguishes between the world as it is in itself and the world as structured by our concepts, insisting that we can know only the latter world, a world we are partly responsible for creating. But unlike Kant, Kuhn thinks that this world changes whenever our scientific theories change. Thus, according to Kuhn, the world that scientists study at a particular time has only a limited lifespan. There is no single, objective notion of truth that can be applied to all such worlds and thus no objective sense in which science can be said to progress by getting closer to the truth. See Peter Lipton, "Coordinating Science," *Nature* 364 (1993): 770. Scientific realists find views such as Kuhn's false and abhorrent. See chapter 9.

10. See editorial footnote, p. 12.

11. Since all the planets, including the earth, circle the sun, the planets that lie inside the earth's orbit, namely Mercury and Venus, will never be seen at more than a certain angular distance from the sun. The planets move more slowly, the further their distance from the sun. Retrograde motion occurs either when the earth is overtaken on the inside by Mercury and Venus or when the earth overtakes the planets (such as Mars, Jupiter, and Saturn—the so-called superior planets) that lie outside of the earth's orbit. Necessarily, then, retrograde motion will occur only when the planet, the earth, and the sun lie in a straight line. For the superior planets, this will be when the planet lies opposite the sun as seen from the earth (what astronomers call opposition); for the inferior planets, retrograde motion will occur only when the planet passes between the sun and the earth (what astronomers call conjunction).

12. See, for example, her criticism of Hempel's satisfaction criterion in chapter 2 of *Science as Social Knowledge*.

13. For a helpful introduction to Habermas's views about science, placing them in the context of thinkers such as Max Weber and Karl Marx, see Robert Hollinger, "From Weber to Habermas," in *Introductory Readings in the Philosophy of Science*, rev. ed., ed. E. D. Klemke, Robert Hollinger, and A. David Kline (Buffalo, N.Y.: Prometheus Books, 1988), 416–26.

14. Trofim Lysenko, a convinced believer in the Lamarckian doctrine of the inheritance of acquired characteristics, obtained dictatorial power over Soviet biology under Stalin. Scientists and teachers who disagreed with Lysenko lost their jobs and some, their lives. This grim episode set Soviet biology back for decades.

15. Longino's own formulation of her fourth criterion is somewhat problematic, since it specifies that intellectual authority in science should be shared among "qualified practitioners." Obviously, if "qualified practitioners" means something like "having received a first-rate scientific training," no violation of the criterion will have occurred if women or racial minorities are denied such training or discouraged from seeking it.

16. Helen E. Longino, "In Search of Feminist Epistemology," *Monist* 77 (1994): 427–85.

17. Barbara McClintock (1902–92) received the Nobel Prize for Physiology or Medicine in 1983 for her pioneering work on “jumping genes” in maize. Using classical genetical and cytological techniques, McClintock concluded that genetic elements in the chromosomes of maize can move around in response to changes in the chemical environment of the cell. Although she was not herself a feminist, McClintock’s research has been regarded as a good illustration of a distinctively feminist approach to science. See, for example, Evelyn Fox Keller, *A Feeling for the Organism: The Life and Work of Barbara McClintock* (San Francisco: W. H. Freeman, 1983). It should be noted that, although McClintock’s theory of transposable elements was ignored for several decades, she was far from being a scientific outsider. In fact, she was elected to the United States Academy of Sciences in 1944 and received many honors and awards for her experimental work in cytogenetics, including the Kimber Genetics Award (1967) and the National Medal of Science (1970). In 1981, she was the first recipient of the MacArthur Laureate Award, granting her a lifetime fellowship of \$60,000 a year, tax free. Keller comments that “if Barbara McClintock’s story illustrates the fallibility of science, it also bears witness to the underlying health of the scientific enterprise. Her eventual vindication demonstrates the capacity of science to overcome its own characteristic kinds of myopia, reminding us that its limitations do not reinforce themselves indefinitely” (Keller, 197). Despite respect for her experimental work, McClintock’s theoretical ideas were hard for many scientists to accept in the 1950s and 1960s because they ran counter to the prevailing paradigm that genes are fixed, stable units of heredity and that information always flows from genes to proteins, never in the reverse direction.



3 |

The Duhem-Quine Thesis and Underdetermination

INTRODUCTION

Probably no set of doctrines has had a greater influence on modern philosophy of science than those included under the designation of the Duhem-Quine thesis. Thinkers as diverse as Sandra Harding, Bas van Fraassen, Mary Hesse, David Bloor, Arthur Fine, Helen Longino, Thomas Kuhn, and Richard Rorty have invoked, in one form or another, a version of the Duhem-Quine thesis to reach conclusions about the limitations of empirical evidence and the rules of scientific method as a constraint on our acceptance or rejection of scientific theories. Some of these philosophers have argued that no scientific theory can ever be conclusively refuted (thus repudiating one of the central features of Popper's philosophy of science); others of them have concluded that, given the presumed gap separating theory from evidence and the presumed insufficiency of methodological rules in bridging that gap, we should never accept any theory as objectively true no matter how well it agrees with the available evidence (thus abandoning scientific realism in favor of some version of skeptical relativism). As we shall see, an astonishing variety of doctrines fall under the Duhem-Quine umbrella. The aim of the readings and commentary in this chapter is to identify these doctrines, to understand how they are related, and to provide a framework for assessing their credibility.

The chapter begins, appropriately enough, with readings from Pierre Duhem and W. V. Quine. The selection by Duhem comes from his book *The Aim and Structure of Physical Theory*, first published (in French) in 1906 (although the papers on which it was based were written in the 1890s). The piece from Quine, "Two Dogmas of Empiricism" (1951), is a modern classic that has had a wide-ranging influence on many areas of philosophy, especially epistemology and the philosophy of science. As Donald Gillies points out in his article, "The Duhem Thesis and the

Quine Thesis," the differences in the scope and focus of the arguments of Duhem and Quine are so great that it makes little sense, historically, to link the names of the two men together as espousing a single, common thesis. Nonetheless, Gillies thinks that, historical considerations aside, we can assemble from the different but related views of Duhem and Quine a plausible version of holism that can aptly be named the Duhem-Quine thesis.

In "Demystifying Underdetermination" Larry Laudan launches a spirited attack on all versions of the underdetermination thesis that have been espoused by Quine and his followers. Laudan is especially critical of those who—like Kuhn, Hesse, and Bloor—have used Duhem-Quine-style arguments to bolster their view that science is governed to a large degree by sociological forces (not logic and scientific method) and can be understood only by taking these historical and social factors into account. Laudan argues that once one distinguishes different versions of the underdetermination thesis, underdetermination shows itself to be either true but innocuous or dramatic and false. All too often, he warns, philosophers take for granted a radical version of the underdetermination thesis without giving anything like a plausible argument to support it.

Physical Theory and Experiment

1 | **The Experimental Testing of a Theory Does Not Have the Same Logical Simplicity in Physics as in Physiology**

The sole purpose of physical theory is to provide a representation and classification of experimental laws; the only test permitting us to judge a physical theory and pronounce it good or bad is the comparison between the consequences of this theory and the experimental laws it has to represent and classify. Now that we have minutely analyzed the characteristics of a physical experiment and of a physical law, we can establish the principles that should govern the comparison between experiment and theory: we can tell how we shall recognize whether a theory is confirmed or weakened by facts.

When many philosophers talk about experimental sciences, they think only of sciences still close to their origins, e.g., physiology or certain branches of chemistry where the experimenter reasons directly on the facts by a method which is only common sense brought to greater attentiveness but where mathematical theory has not yet introduced its symbolic representations. In such sciences the comparison between the deductions of a theory and the facts of experiment is subject to very simple rules. These rules were formulated in a particularly forceful manner by Claude Bernard, who would condense them into a single principle, as follows:

"The experimenter should suspect and stay away from fixed ideas, and always preserve his freedom of mind.

"The first condition that has to be fulfilled by a scientist who is devoted to the investigation of natural phenomena is to preserve a complete freedom of mind based on philosophical doubt."

If a theory suggests experiments to be done, so much the better: "... we can follow our judgment and our thought, give free rein to our imagination provided that all our ideas are only pretexts for instituting new experiments that may furnish us probative facts or unexpected and fruitful ones."² Once the experiment is done and the results clearly established, if a theory takes them over in order to generalize them, coordinate them, and draw from them new subjects for experiment, still so much the better: "... if one is imbued with the principles of experimental method, there is nothing to fear; for so long as the idea is a right one, it will go on being developed; when it is an erroneous idea, experiment is there to correct it."³ But so long as the experiment lasts, the theory should remain waiting, under strict orders to stay outside the door of the laboratory; it should keep silent and leave the scientist without disturbing him while he faces the facts directly; the facts must be observed without a preconceived idea and gathered with the same scrupulous impartiality, whether they confirm or contradict the predictions of the theory. The report that the observer will give us of his experiment should be a faithful and scrupulously exact reproduction of the phenomena, and should not let us even guess what system the scientist places his confidence in or distrusts.

"Men who have an excessive faith in their theories or in their ideas are not only poorly disposed to make discoveries but they also make very poor observations. They necessarily observe with a preconceived idea and, when they have begun an experiment, they want to see in its results only a confirmation of their theory. Thus they distort observation and often neglect very important facts because they go counter to their goal. That is what made us say elsewhere that we must never do experiments in order to confirm our ideas but merely to check them. . . . But it quite naturally happens that those who believe too much in their own theories do not sufficiently believe in the theories of others. Then the dominant idea of these condemners of others is to find fault with the theories of the latter and to seek to contradict them. The setback for science remains the same. They are doing experiments only in order to destroy a theory instead of doing them in order to look for the truth. They also make poor observations because they take into the results of their experiments only what fits their purpose, by neglecting what is unrelated to it, and by very carefully avoiding whatever might go in the direction of the idea they wish to combat. Thus one is led by two parallel paths to the same result, that is to say, to falsifying science and the facts.

"The conclusion of all this is that it is necessary to obliterate one's opinion as well as that of others when faced with the decisions of the experiment; . . . we must accept the results of experiment just as they present themselves with all that is unforeseen and accidental in them."⁴

Here, for example, is a physiologist who admits that the anterior roots of the spinal nerve contain the motor nerve-fibers and the posterior roots the sensory fibers. The theory he accepts leads him to imagine an exper-

iment: if he cuts a certain anterior root, he ought to be suppressing the mobility of a certain part of the body without destroying its sensibility; after making the section of this root, when he observes the consequences of his operation and when he makes a report of it, he must put aside all his ideas concerning the physiology of the spinal nerve; his report must be a raw description of the facts; he is not permitted to overlook or fail to mention any movement or quiver contrary to his predictions or to attribute it to some secondary cause unless some special experiment has given evidence of this cause; he must, if he does not wish to be accused of scientific bad faith, establish an absolute separation or watertight compartment between the consequences of his theoretical deductions and the establishing of the facts shown by his experiments.

Such a rule is not by any means easily followed; it requires of the scientist an absolute detachment from his own thought and a complete absence of animosity when confronted with the opinion of another person; neither vanity nor envy ought to be countenanced by him. As Bacon put it, he should never show eyes lustrous with human passions. Freedom of mind, which constitutes the sole principle of experimental method, according to Claude Bernard, does not depend merely on intellectual conditions, but also on moral conditions, making its practice rarer and more meritorious.

But if experimental method as just described is difficult to practice, the logical analysis of it is very simple. This is no longer the case when the theory to be subjected to test by the facts is not a theory of physiology but a theory of physics. In the latter case, in fact, it is impossible to leave outside the laboratory door the theory that we wish to test, for without theory it is impossible to regulate a single instrument or to interpret a single reading. We have seen that in the mind of the physicist there are constantly present two sorts of apparatus: one is the concrete apparatus in glass and metal, manipulated by him, the other is the schematic and abstract apparatus which theory substitutes for the concrete apparatus and on which the physicist does his reasoning. For these two ideas are indissolubly connected in his intelligence, and each necessarily calls on the other; the physicist can no sooner conceive the concrete apparatus without associating with it the idea of the schematic apparatus than a Frenchman can conceive an idea without associating it with the French word expressing it. This radical impossibility, preventing one from dissociating physical theories from the experimental procedures appropriate for testing these theories, complicates this test in a singular way, and obliges us to examine the logical meaning of it carefully.

Of course, the physicist is not the only one who appeals to theories at the very time he is experimenting or reporting the results of his experiments. The chemist and the physiologist when they make use of physical instruments, e.g., the thermometer, the manometer, the calorimeter, the galvanometer, and the saccharimeter, implicitly admit the accuracy of

the theories justifying the use of these pieces of apparatus as well as of the theories giving meaning to the abstract ideas of temperature, pressure, quantity of heat, intensity of current, and polarized light, by means of which the concrete indications of these instruments are translated. But the theories used, as well as the instruments employed, belong to the domain of physics; by accepting with these instruments the theories without which their readings would be devoid of meaning, the chemist and the physiologist show their confidence in the physicist, whom they suppose to be infallible. The physicist, on the other hand, is obliged to trust his own theoretical ideas or those of his fellow-physicists. From the standpoint of logic, the difference is of little importance; for the physiologist and chemist as well as for the physicist, the statement of the result of an experiment implies, in general, an act of faith in a whole group of theories.

2 | An Experiment in Physics Can Never Condemn an Isolated Hypothesis but Only a Whole Theoretical Group

The physicist who carries out an experiment, or gives a report of one, implicitly recognizes the accuracy of a whole group of theories. Let us accept this principle and see what consequences we may deduce from it when we seek to estimate the role and logical import of a physical experiment.

In order to avoid any confusion we shall distinguish two sorts of experiments: experiments of *application*, which we shall first just mention, and experiments of *testing*, which will be our chief concern.

You are confronted with a problem in physics to be solved practically; in order to produce a certain effect you wish to make use of knowledge acquired by physicists; you wish to light an incandescent bulb; accepted theories indicate to you the means for solving the problem; but to make use of these means you have to secure certain information; you ought, I suppose, to determine the electromotive force of the battery of generators at your disposal; you measure this electromotive force: that is what I call an experiment of application. This experiment does not aim at discovering whether accepted theories are accurate or not; it merely intends to draw on these theories. In order to carry it out, you make use of instruments that these same theories legitimize; there is nothing to shock logic in this procedure.

But experiments of application are not the only ones the physicist has to perform; only with their aid can science aid practice, but it is not through them that science creates and develops itself; besides experiments of application, we have experiments of testing.

A physicist disputes a certain law; he calls into doubt a certain theo-

retical point. How will he justify these doubts? How will he demonstrate the inaccuracy of the law? From the proposition under indictment he will derive the prediction of an experimental fact; he will bring into existence the conditions under which this fact should be produced; if the predicted fact is not produced, the proposition which served as the basis of the prediction will be irremediably condemned.

F. E. Neumann assumed that in a ray of polarized light the vibration is parallel to the plane of polarization, and many physicists have doubted this proposition. How did O. Wiener undertake to transform this doubt into a certainty in order to condemn Neumann's proposition? He deduced from this proposition the following consequence: If we cause a light beam reflected at 45° from a plate of glass to interfere with the incident beam polarized perpendicularly to the plane of incidence, there ought to appear alternately dark and light interference bands parallel to the reflecting surface; he brought about the conditions under which these bands should have been produced and showed that the predicted phenomenon did not appear, from which he concluded that Neumann's proposition is false, viz., that in a polarized ray of light the vibration is not parallel to the plane of polarization.

Such a mode of demonstration seems as convincing and as irrefutable as the proof by reduction to absurdity customary among mathematicians; moreover, this demonstration is copied from the reduction to absurdity, experimental contradiction playing the same role in one as logical contradiction plays in the other.

Indeed, the demonstrative value of experimental method is far from being so rigorous or absolute: the conditions under which it functions are much more complicated than is supposed in what we have just said; the evaluation of results is much more delicate and subject to caution.

A physicist decides to demonstrate the inaccuracy of a proposition: in order to deduce from this proposition the prediction of a phenomenon and institute the experiment which is to show whether this phenomenon is or is not produced, in order to interpret the results of this experiment and establish that the predicted phenomenon is not produced, he does not confine himself to making use of the proposition in question; he makes use also of a whole group of theories accepted by him as beyond dispute. The prediction of the phenomenon, whose nonproduction is to cut off debate, does not derive from the proposition challenged if taken by itself, but from the proposition at issue joined to that whole group of theories; if the predicted phenomenon is not produced, not only is the proposition questioned at fault, but so is the whole theoretical scaffolding used by the physicist. The only thing the experiment teaches us is that among the propositions used to predict the phenomenon and to establish whether it would be produced, there is at least one error; but where this error lies is just what it does not tell us. The physicist may declare that this error is contained in exactly the proposition he wishes to refute, but is he sure it

is not in another proposition? If he is, he accepts implicitly the accuracy of all the other propositions he has used, and the validity of his conclusion is as great as the validity of his confidence.

Let us take as an example the experiment imagined by Zenker and carried out by O. Wiener. In order to predict the formation of bands in certain circumstances and to show that these did not appear, Wiener did not make use merely of the famous proposition of F. E. Neumann, the proposition which he wished to refute; he did not merely admit that in a polarized ray vibrations are parallel to the plane of polarization; but he used, besides this, propositions, laws, and hypotheses constituting the optics commonly accepted: he admitted that light consists in simple periodic vibrations, that these vibrations are normal to the light ray, that at each point the mean kinetic energy of the vibratory motion is a measure of the intensity of light, that the more or less complete attack of the gelatine coating on a photographic plate indicates the various degrees of this intensity. By joining these propositions, and many others that would take too long to enumerate, to Neumann's proposition, Wiener was able to formulate a forecast and establish that the experiment belied it. If he attributed this solely to Neumann's proposition, if it alone bears the responsibility for the error this negative result has put in evidence, then Wiener was taking all the other propositions he invoked as beyond doubt. But this assurance is not imposed as a matter of logical necessity; nothing stops us from taking Neumann's proposition as accurate and shifting the weight of the experimental contradiction to some other proposition of the commonly accepted optics; as H. Poincaré has shown, we can very easily rescue Neumann's hypothesis from the grip of Wiener's experiment on the condition that we abandon in exchange the hypothesis which takes the mean kinetic energy as the measure of the light intensity; we may, without being contradicted by the experiment, let the vibration be parallel to the plane of polarization, provided that we measure the light intensity by the mean potential energy of the medium deforming the vibratory motion.

These principles are so important that it will be useful to apply them to another example; again we choose an experiment regarded as one of the most decisive ones in optics.

We know that Newton conceived the emission theory for optical phenomena. The emission theory supposes light to be formed of extremely thin projectiles, thrown out with very great speed by the sun and other sources of light; these projectiles penetrate all transparent bodies; on account of the various parts of the media through which they move, they undergo attractions and repulsions; when the distance separating the acting particles is very small these actions are very powerful, and they vanish when the masses between which they act are appreciably far from each other. These essential hypotheses joined to several others, which we pass over without mention, lead to the formulation of a complete theory of reflection and refraction of light; in particular, they imply the following

proposition: The index of refraction of light passing from one medium into another is equal to the velocity of the light projectile within the medium it penetrates, divided by the velocity of the same projectile in the medium it leaves behind.

This is the proposition that Arago chose in order to show that the theory of emission is in contradiction with the facts. From this proposition a second follows: Light travels faster in water than in air. Now Arago had indicated an appropriate procedure for comparing the velocity of light in air with the velocity of light in water; the procedure, it is true, was inapplicable, but Foucault modified the experiment in such a way that it could be carried out; he found that the light was propagated less rapidly in water than in air. We may conclude from this, with Foucault, that the system of emission is incompatible with the facts.

I say the *system* of emission and not the *hypothesis* of emission; in fact, what the experiment declares stained with error is the whole group of propositions accepted by Newton, and after him by Laplace and Biot, that is, the whole theory from which we deduce the relation between the index of refraction and the velocity of light in various media. But in condemning this system as a whole by declaring it stained with error, the experiment does not tell us where the error lies. Is it in the fundamental hypothesis that light consists in projectiles thrown out with great speed by luminous bodies? Is it in some other assumption concerning the actions experienced by light corpuscles due to the media through which they move? We know nothing about that. It would be rash to believe, as Arago seems to have thought, that Foucault's experiment condemns once and for all the very hypothesis of emission, i.e., the assimilation of a ray of light to a swarm of projectiles. If physicists had attached some value to this task, they would undoubtedly have succeeded in founding on this assumption a system of optics that would agree with Foucault's experiment.

In sum, the physicist can never subject an isolated hypothesis to experimental test, but only a whole group of hypotheses; when the experiment is in disagreement with his predictions, what he learns is that at least one of the hypotheses constituting this group is unacceptable and ought to be modified; but the experiment does not designate which one should be changed.

We have gone a long way from the conception of the experimental method arbitrarily held by persons unfamiliar with its actual functioning. People generally think that each one of the hypotheses employed in physics can be taken in isolation, checked by experiment, and then, when many varied tests have established its validity, given a definitive place in the system of physics. In reality, this is not the case. Physics is not a machine which lets itself be taken apart; we cannot try each piece in isolation and, in order to adjust it, wait until its solidity has been carefully checked. Physical science is a system that must be taken as a whole; it is

an organism in which one part cannot be made to function except when the parts that are most remote from it are called into play, some more so than others, but all to some degree. If something goes wrong, if some discomfort is felt in the functioning of the organism, the physicist will have to ferret out through its effect on the entire system which organ needs to be remedied or modified without the possibility of isolating this organ and examining it apart. The watchmaker to whom you give a watch that has stopped separates all the wheelworks and examines them one by one until he finds the part that is defective or broken. The doctor to whom a patient appears cannot dissect him in order to establish his diagnosis; he has to guess the seat and cause of the ailment solely by inspecting disorders affecting the whole body. Now, the physicist concerned with remedying a limping theory resembles the doctor and not the watchmaker.

3 | A "Crucial Experiment" Is Impossible in Physics

Let us press this point further, for we are touching on one of the essential features of experimental method, as it is employed in physics.

Reduction to absurdity seems to be merely a means of refutation, but it may become a method of demonstration: in order to demonstrate the truth of a proposition it suffices to corner anyone who would admit the contradictory of the given proposition into admitting an absurd consequence. We know to what extent the Greek geometers drew heavily on this mode of demonstration.

Those who assimilate experimental contradiction to reduction to absurdity imagine that in physics we may use a line of argument similar to the one Euclid employed so frequently in geometry. Do you wish to obtain from a group of phenomena a theoretically certain and indisputable explanation? Enumerate all the hypotheses that can be made to account for this group of phenomena; then, by experimental contradiction eliminate all except one; the latter will no longer be a hypothesis, but will become a certainty.

Suppose, for instance, we are confronted with only two hypotheses. Seek experimental conditions such that one of the hypotheses forecasts the production of one phenomenon and the other the production of quite a different effect; bring these conditions into existence and observe what happens; depending on whether you observe the first or the second of the predicted phenomena, you will condemn the second or the first hypothesis; the hypothesis not condemned will be henceforth indisputable; debate will be cut off, and a new truth will be acquired by science. Such is the experimental test that the author of the *Novum Organum* [Francis Bacon] called the "*fact of the cross*," borrowing this expression from the crosses which at an intersection indicate the various roads.

We are confronted with two hypotheses concerning the nature of light; for Newton, Laplace, or Biot light consisted of projectiles hurled with extreme speed, but for Huygens, Young, or Fresnel light consisted of vibrations whose waves are propagated within an ether. These are the only two possible hypotheses as far as one can see: either the motion is carried away by the body it excites and remains attached to it, or else it passes from one body to another. Let us pursue the first hypothesis; it declares that light travels more quickly in water than in air; but if we follow the second, it declares that light travels more quickly in air than in water. Let us set up Foucault's apparatus; we set into motion the turning mirror; we see two luminous spots formed before us, one colorless, the other greenish. If the greenish band is to the left of the colorless one, it means that light travels faster in water than in air, and that the hypothesis of vibrating waves is false. If, on the contrary, the greenish band is to the right of the colorless one, that means that light travels faster in air than in water, and that the hypothesis of emissions is condemned. We look through the magnifying glass used to examine the two luminous spots, and we notice that the greenish spot is to the right of the colorless one; the debate is over; light is not a body, but a vibratory wave motion propagated by the ether; the emission hypothesis has had its day; the wave hypothesis has been put beyond doubt, and the crucial experiment has made it a new article of the scientific credo.

What we have said in the foregoing paragraph shows how mistaken we should be to attribute to Foucault's experiment so simple a meaning and so decisive an importance; for it is not between two hypotheses, the emission and wave hypotheses, that Foucault's experiment judges trenchantly; it decides rather between two sets of theories each of which has to be taken as a whole, i.e., between two entire systems, Newton's optics and Huygens' optics.

But let us admit for a moment that in each of these systems everything is compelled to be necessary by strict logic, except a single hypothesis; consequently, let us admit that the facts, in condemning one of the two systems, condemn once and for all the single doubtful assumption it contains. Does it follow that we can find in the "crucial experiment" an irrefutable procedure for transforming one of the two hypotheses before us into a demonstrated truth? Between two contradictory theorems of geometry there is no room for a third judgment; if one is false, the other is necessarily true. Do two hypotheses in physics ever constitute such a strict dilemma? Shall we ever dare to assert that no other hypothesis is imaginable? Light may be a swarm of projectiles, or it may be a vibratory motion whose waves are propagated in a medium; is it forbidden to be anything else at all? Arago undoubtedly thought so when he formulated this incisive alternative: Does light move more quickly in water than in air? "Light is a body. If the contrary is the case, then light is a wave." But it would be difficult for us to take such a decisive stand; Maxwell, in fact, showed that

we might just as well attribute light to a periodical electrical disturbance that is propagated within a dielectric medium.

Unlike the reduction to absurdity employed by geometers, experimental contradiction does not have the power to transform a physical hypothesis into an indisputable truth; in order to confer this power on it, it would be necessary to enumerate completely the various hypotheses which may cover a determinate group of phenomena; but the physicist is never sure he has exhausted all the imaginable assumptions. The truth of a physical theory is not decided by heads or tails.

4 | Criticism of the Newtonian Method. First Example: Celestial Mechanics

It is illusory to seek to construct by means of experimental contradiction a line of argument in imitation of the reduction to absurdity; but the geometer is acquainted with other methods for attaining certainty than the method of reducing to an absurdity; the direct demonstration in which the truth of a proposition is established by itself and not by the refutation of the contradictory proposition seems to him the most perfect of arguments. Perhaps physical theory would be more fortunate in its attempts if it sought to imitate direct demonstration. The hypotheses from which it starts and develops its conclusions would then be tested one by one; none would have to be accepted until it presented all the certainty that experimental method can confer on an abstract and general proposition; that is to say, each would necessarily be either a law drawn from observation by the sole use of those two intellectual operations called induction and generalization, or else a corollary mathematically deduced from such laws. A theory based on such hypotheses would then not present anything arbitrary or doubtful; it would deserve all the confidence merited by the faculties which serve us in formulating natural laws.

It was this sort of physical theory that Newton had in mind when, in the "General Scholium" which crowns his *Principia*, he rejected so vigorously as outside of natural philosophy any hypothesis that induction did not extract from experiment; when he asserted that in a sound physics every proposition should be drawn from phenomena and generalized by induction.

The ideal method we have just described therefore deserves to be named the Newtonian method. Besides, did not Newton follow this method when he established the system of universal attraction, thus adding to his precepts the most magnificent of examples? Is not his theory of gravitation derived entirely from the laws which were revealed to Kepler by observation, laws which problematic reasoning transforms and whose consequences induction generalizes?

This first law of Kepler's, "The radial vector from the sun to a planet sweeps out an area proportional to the time during which the planet's motion is observed," did, in fact, teach Newton that each planet is constantly subjected to a force directed toward the sun.

The second law of Kepler's, "The orbit of each planet is an ellipse having the sun at one focus," taught him that the force attracting a given planet varies with the distance of this planet from the sun, and that it is in an inverse ratio to the square of this distance.

The third law of Kepler's, "The squares of the periods of revolution of the various planets are proportional to the cubes of the major axes of their orbits," showed him that different planets would, if they were brought to the same distance from the sun, undergo in relation to its attractions proportional to their respective masses.

The experimental laws established by Kepler and transformed by geometric reasoning yield all the characteristics present in the action exerted by the sun on a planet; by induction Newton generalized the result obtained; he allowed this result to express the law according to which any portion of matter acts on any other portion whatsoever, and he formulated this great principle: "Any two bodies whatsoever attract each other with a force which is proportional to the product of their masses and in inverse ratio to the square of the distance between them." The principle of universal gravitation was found, and it was obtained, without any use having been made of any fictive hypothesis, by the inductive method the plan of which Newton outlined.

Let us again examine this application of the Newtonian method, this time more closely; let us see if a somewhat strict logical analysis will leave intact the appearance of rigor and simplicity that this very summary exposition attributes to it.

In order to assure this discussion of all the clarity it needs, let us begin by recalling the following principle, familiar to all those who deal with mechanics: We cannot speak of the force which attracts a body in given circumstances before we have designated the supposedly fixed term of reference to which we relate the motion of all bodies; when we change this point of reference or term of comparison, the force representing the effect produced on the observed body by the other bodies surrounding it changes in direction and magnitude according to the rules stated by mechanics with precision.

That posited, let us follow Newton's reasoning.

Newton first took the sun as the fixed point of reference; he considered the motions affecting the different planets by reference to the sun; he admitted Kepler's laws as governing these motions, and derived the following proposition: If the sun is the point of reference in relation to which all forces are compared, each planet is subjected to a force directed toward the sun, a force proportional to the mass of the planet and to the

inverse square of its distance from the sun. Since the latter is taken as the reference point, it is not subject to any force.

In an analogous manner Newton studied the motion of the satellites and for each of these he chose as a fixed reference point the planet which the satellite accompanies, the earth in the case of the moon, Jupiter in the case of the masses moving around Jupiter. Laws just like Kepler's were taken as governing these motions, from which it follows that we can formulate the following proposition: If we take as a fixed reference point the planet accompanied by a satellite, this satellite is subject to a force directed toward the planet varying inversely with the square of the distance. If, as happens with Jupiter, the same planet possesses several satellites, these satellites, were they at the same distance from the planet, would be acted on by the latter with forces proportional to their respective masses. The planet is itself not acted on by the satellite.

Such, in very precise form, are the propositions which Kepler's laws of planetary motion and the extension of these laws to the motions of satellites authorize us to formulate. For these propositions Newton substituted another which may be stated as follows: Any two celestial bodies whatsoever exert on each other a force of attraction in the direction of the straight line joining them, a force proportional to the product of their masses and to the inverse square of the distance between them. This statement presupposes all motions and forces to be related to the same reference point; the latter is an ideal standard of reference which may well be conceived by the geometer but which does not characterize in an exact and concrete manner the position in the sky of any body.

Is this principle of universal gravitation merely a generalization of the two statements provided by Kepler's laws and their extension to the motion of satellites? Can induction derive it from these two statements? Not at all. In fact, not only is it more general than these two statements and unlike them, but it contradicts them. The student of mechanics who accepts the principle of universal attraction can calculate the magnitude and direction of the forces between the various planets and the sun when the latter is taken as the reference point, and if he does he finds that these forces are not what our first statement would require. He can determine the magnitude and direction of each of the forces between Jupiter and its satellites when we refer all the motions to the planet, assumed to be fixed, and if he does he notices that these forces are not what our second statement would require.

The principle of universal gravity, very far from being derivable by generalization and induction from the observational laws of Kepler, formally contradicts these laws. If Newton's theory is correct, Kepler's laws are necessarily false.

Kepler's laws based on the observation of celestial motions do not transfer their immediate experimental certainty to the principle of univer-

sal weight, since if, on the contrary, we admit the absolute exactness of Kepler's laws, we are compelled to reject the proposition on which Newton based his celestial mechanics. Far from adhering to Kepler's laws, the physicist who claims to justify the theory of universal gravitation finds that he has, first of all, to resolve a difficulty in these laws: he has to prove that his theory, incompatible with the exactness of Kepler's laws, subjects the motions of the planets and satellites to other laws scarcely different enough from the first laws for Tycho Brahé, Kepler, and their contemporaries to have been able to discern the deviations between the Keplerian and Newtonian orbits. This proof derives from the circumstances that the sun's mass is very large in relation to the masses of the various planets and the mass of a planet is very large in relation to the masses of its satellites.

Therefore, if the certainty of Newton's theory does not emanate from the certainty of Kepler's laws, how will this theory prove its validity? It will calculate, with all the high degree of approximation that the constantly perfected methods of algebra involve, the perturbations which at each instant remove every heavenly body from the orbit assigned to it by Kepler's laws; then it will compare the calculated perturbations with the perturbations observed by means of the most precise instruments and the most scrupulous methods. Such a comparison will not only bear on this or that part of the Newtonian principle, but will involve all its parts at the same time; with those it will also involve all the principles of dynamics; besides, it will call in the aid of all the propositions of optics, the statics of gases, and the theory of heat, which are necessary to justify the properties of telescopes in their construction, regulation, and correction, and in the elimination of the errors caused by diurnal or annual aberration and by atmospheric refraction. It is no longer a matter of taking, one by one, laws justified by observation, and raising each of them by induction and generalization to the rank of a principle; it is a matter of comparing the corollaries of a whole group of hypotheses to a whole group of facts.

Now, if we seek out the causes which have made the Newtonian method fail in this case for which it was imagined and which seemed to be the most perfect application for it, we shall find them in that double character of any law made use of by theoretical physics: This law is symbolic and approximate.

Undoubtedly, Kepler's laws bear quite directly on the very objects of astronomical observation; they are as little symbolic as possible. But in this purely experimental form they remain inappropriate for suggesting the principle of universal gravitation; in order to acquire this fecundity they must be transformed and must yield the characters of the forces by which the sun attracts the various planets.

Now this new form of Kepler's laws is a symbolic form; only dynamics gives meanings to the words "force" and "mass," which serve to state it, and only dynamics permits us to substitute the new symbolic formulas for

the old realistic formulas, to substitute statements relative to "forces" and "masses" for laws relative to orbits. The legitimacy of such a substitution implies full confidence in the laws of dynamics.

And in order to justify this confidence let us not proceed to claim that the laws of dynamics were beyond doubt at the time Newton made use of them in symbolically translating Kepler's laws; that they had received enough empirical confirmation to warrant the support of reason. In fact, the laws of dynamics had been subjected up to that time to only very limited and very crude tests. Even their enunciations had remained very vague and involved; only in Newton's *Principia* had they been for the first time formulated in a precise manner. It was in the agreement of the facts with the celestial mechanics which Newton's labors gave birth to that they received their first convincing verification.

Thus the translation of Kepler's laws into symbolic laws, the only kind useful for a theory, presupposed the prior adherence of the physicist to a whole group of hypotheses. But, in addition, Kepler's laws being only approximate laws, dynamics permitted giving them an infinity of different symbolic translations. Among these various forms, infinite in number, there is one and only one which agrees with Newton's principle. The observations of Tycho Brahé, so felicitously reduced to laws by Kepler, permit the theorist to choose this form, but they do not constrain him to do so, for there is an infinity of others they permit him to choose.

The theorist cannot, therefore, be content to invoke Kepler's laws in order to justify his choice. If he wishes to prove that the principle he has adopted is truly a principle of natural classification for celestial motions, he must show that the observed perturbations are in agreement with those which had been calculated in advance; he has to show how from the course of Uranus he can deduce the existence and position of a new planet, and find Neptune in an assigned direction at the end of his telescope. . . .

8 | Are Certain Postulates of Physical Theory Incapable of Being Refuted by Experiment?

We recognize a correct principle by the facility with which it straightens out the complicated difficulties into which the use of erroneous principles brought us.

If, therefore, the idea we have put forth is correct, namely, that comparison is established necessarily between the *whole* of theory and the *whole* of experimental facts, we ought in the light of this principle to see the disappearance of the obscurities in which we should be lost by thinking that we are subjecting each isolated theoretical hypothesis to the test of facts.

Foremost among the assertions in which we shall aim at eliminating the appearance of paradox, we shall place one that has recently been often formulated and discussed. Stated first by G. Milhaud in connection with the "*pure bodies*" of chemistry,⁵ it has been developed at length and forcefully by H. Poincaré with regard to principles of mechanics;⁶ Edouard Le Roy has also formulated it with great clarity.⁷

That assertion is as follows: Certain fundamental hypotheses of physical theory cannot be contradicted by any experiment, because they constitute in reality *definitions*, and because certain expressions in the physicist's usage take their meaning only through them.

Let us take one of the examples cited by Le Roy:

When a heavy body falls freely, the acceleration of its fall is constant. Can such a law be contradicted by experiment? No, for it constitutes the very definition of what is meant by "falling freely." If while studying the fall of a heavy body we found that this body does not fall with uniform acceleration, we should conclude not that the stated law is false, but that the body does not fall freely, that some cause obstructs its motion, and that the deviations of the observed facts from the law as stated would serve to discover this cause and to analyze its effects.

Thus, M. Le Roy concludes, "laws are verifiable, taking things strictly . . . , because they constitute the very criterion by which we judge appearances as well as the methods that it would be necessary to utilize in order to submit them to an inquiry whose precision is capable of exceeding any assignable limit."

Let us study again in greater detail, in the light of the principles previously set down, what this comparison is between the law of falling bodies and experiment.

Our daily observations have made us acquainted with a whole category of motions which we have brought together under the name of motions of heavy bodies; among these motions is the falling of a heavy body when it is not hindered by any obstacle. The result of this is that the words "free fall of a heavy body" have a meaning for the man who appeals only to the knowledge of common sense and who has no notion of physical theories.

On the other hand, in order to classify the laws of motion in question the physicist has created a theory, the theory of weight, an important application of rational mechanics. In that theory, intended to furnish a symbolic representation of reality, there is also the question of "free fall of a heavy body," and as a consequence of the hypotheses supporting this whole scheme free fall must necessarily be a uniformly accelerated motion.

The words "free fall of a heavy body" now have two distinct meanings. For the man ignorant of physical theories, they have their *real* meaning, and they mean what common sense means in pronouncing them; for the physicist they have a *symbolic* meaning, and mean "uniformly accelerated

motion." Theory would not have realized its aim if the second meaning were not the sign of the first, if a fall regarded as free by common sense were not also regarded as uniformly accelerated, or nearly uniformly accelerated, since common-sense observations are essentially devoid of precision, according to what we have already said.

This agreement, without which the theory would have been rejected without further examination, is finally arrived at: a fall declared by common sense to be nearly free is also a fall whose acceleration is nearly constant. But noticing this crudely approximate agreement does not satisfy us; we wish to push on and surpass the degree of precision which common sense can claim. With the aid of the theory that we have imagined, we put together apparatus enabling us to recognize with sensitive accuracy whether the fall of a body is or is not uniformly accelerated; this apparatus shows us that a certain fall regarded by common sense as a free fall has a slightly variable acceleration. The proposition which in our theory gives its symbolic meaning to the words "free fall" does not represent with sufficient accuracy the properties of the real and concrete fall that we have observed.

Two alternatives are then open to us.

In the first place, we can declare that we were right in regarding the fall studied as a free fall and in requiring that the theoretical definition of these words agree with our observations. In this case, since our theoretical definition does not satisfy this requirement, it must be rejected; we must construct another mechanics on new hypotheses, a mechanics in which the words "free fall" no longer signify "uniformly accelerated motion," but "fall whose acceleration varies according to a certain law."

In the second alternative, we may declare that we were wrong in establishing a connection between the concrete fall we have observed and the symbolic free fall defined by our theory, that the latter was too simplified a scheme of the former, that in order to represent suitably the fall as our experiments have reported it the theorist should give up imagining a weight falling freely and think in terms of a weight hindered by certain obstacles like the resistance of the air, that in picturing the action of these obstacles by means of appropriate hypotheses he will compose a more complicated scheme than a free weight but one more apt to reproduce the details of the experiment; in short, . . . we may seek to eliminate by means of suitable "corrections" the "causes of error," such as air resistance, which influenced our experiment.

M. Le Roy asserts that we shall prefer the second to the first alternative, and he is surely right in this. The reasons dictating this choice are easy to perceive. By taking the first alternative we should be obliged to destroy from top to bottom a very vast theoretical system which represents in a most satisfactory manner a very extensive and complex set of experimental laws. The second alternative, on the other hand, does not make us lose anything of the terrain already conquered by physical theory; in

addition, it has succeeded in so large a number of cases that we can bank with interest on a new success. But in this confidence accorded the law of fall of weights, we see nothing analogous to the certainty that a mathematical definition draws from its very essence, that is, to the kind of certainty we have when it would be foolish to doubt that the various points on a circumference are all equidistant from the center.

We have here nothing more than a particular application of the principle set down in Section 2 of this chapter. A disagreement between the concrete facts constituting an experiment and the symbolic representation which theory substitutes for this experiment proves that some part of this symbol is to be rejected. But which part? This the experiment does not tell us; it leaves to our sagacity the burden of guessing. Now among the theoretical elements entering into the composition of this symbol there is always a certain number which the physicists of a certain epoch agree in accepting without test and which they regard as beyond dispute. Hence, the physicist who wishes to modify this symbol will surely bring his modification to bear on elements other than those just mentioned.

But what impels the physicist to act thus is not logical necessity. It would be awkward and ill inspired for him to do otherwise, but it would not be doing something logically absurd; he would not for all that be walking in the footsteps of the mathematician mad enough to contradict his own definitions. More than this, perhaps some day by acting differently, by refusing to invoke causes of error and take recourse to corrections in order to reestablish agreement between the theoretical scheme and the fact, and by resolutely carrying out a reform among the propositions declared untouchable by common consent, he will accomplish the work of a genius who opens a new career for a theory.

Indeed, we must really guard ourselves against believing forever warranted those hypotheses which have become universally adopted conventions, and whose certainty seems to break through experimental contradiction by throwing the latter back on more doubtful assumptions. The history of physics shows us that very often the human mind has been led to overthrow such principles completely, though they have been regarded by common consent for centuries as inviolable axioms, and to rebuild its physical theories on new hypotheses.

Was there, for instance, a clearer or more certain principle for thousands of years than this one: In a homogeneous medium, light is propagated in a straight line? Not only did this hypothesis carry all former optics, catoptrics, and dioptrics, whose elegant geometric deductions represented at will an enormous number of facts, but it had become, so to speak, the physical definition of a straight line. It is to this hypothesis that any man wishing to make a straight line appeals, the carpenter who verifies the straightness of a piece of wood, the surveyor who lines up his sights, the geodetic surveyor who obtains a direction with the help of the pinholes of his alidade, the astronomer who defines the position of stars by the

optical axis of his telescope. However, the day came when physicists tired of attributing to some cause of error the diffraction effects observed by Grimaldi, when they resolved to reject the law of the rectilinear propagation of light and to give optics entirely new foundations; and this bold resolution was the signal of remarkable progress for physical theory.

9 | On Hypotheses Whose Statement Has No Experimental Meaning

This example, as well as others we could add from the history of science, should show that it would be very imprudent for us to say concerning a hypothesis commonly accepted today: "We are certain that we shall never be led to abandon it because of a new experiment, no matter how precise it is." Yet M. Poincaré does not hesitate to enunciate it concerning the principles of mechanics.*

To the reasons already given to prove that these principles cannot be reached by experimental refutation, M. Poincaré adds one which seems even more convincing: Not only can these principles not be refuted by experiment because they are the universally accepted rules serving to discover in our theories the weak spots indicated by these refutations, but also, they cannot be refuted by experiment because *the operation which would claim to compare them with the facts would have no meaning.*

Let us explain that by an illustration.

The principle of inertia teaches us that a material point removed from the action of any other body moves in a straight line with uniform motion. Now, we can observe only relative motions; we cannot, therefore, give an experimental meaning to this principle unless we assume a certain point chosen or a certain geometric solid taken as a fixed reference point to which the motion of the material point is related. The fixation of this reference frame constitutes an integral part of the statement of the law, for if we omitted it, this statement would be devoid of meaning. There are as many different laws as there are distinct frames of reference. We shall be stating one law of inertia when we say that the motion of an isolated point assumed to be seen from the earth is rectilinear and uniform, and another when we repeat the same sentence in referring the motion to the sun, and still another if the frame of reference chosen is the totality of fixed stars. But then, one thing is indeed certain, namely, that whatever the motion of a material point is, when seen from a first frame of reference, we can always and in infinite ways choose a second frame of reference such that seen from the latter our material point appears to move in a straight line with uniform motion. We cannot, therefore, attempt an experimental verification of the principle of inertia; false when we refer the motions to one frame of reference, it will become true when

selection is made of another term of comparison, and we shall always be free to choose the latter. If the ~~law of inertia stated by taking the earth as~~ a frame of reference is contradicted by an observation, we shall substitute for it the law of inertia whose statement refers the motion to the sun; if the latter in its turn is contraverted, we shall replace the sun in the statement of the law by the system of fixed stars, and so forth. It is impossible to stop this loophole.

The principle of the equality of action and reaction, analyzed at length by M. Poincaré,⁹ provides room for analogous remarks. This principle may be stated thus: "The center of gravity of an isolated system can have only a uniform rectilinear motion."

This is the principle that we propose to verify by experiment. "Can we make this verification? For that it would be necessary for isolated systems to exist. Now, these systems do not exist; the only isolated system is the whole universe.

"But we can observe only relative motions; the absolute motion of the center of the universe will therefore be forever unknown. We shall never be able to know if it is rectilinear and uniform or, better still, the question has no meaning. Whatever facts we may observe, we shall hence always be free to assume our principle is true."

Thus many a principle of mechanics has a form such that it is absurd to ask one's self: "Is this principle in agreement with experiment or not?" This strange character is not peculiar to the principles of mechanics; it also marks certain fundamental hypotheses of our physical or chemical theories.¹⁰

For example, chemical theory rests entirely on the "law of multiple proportions"; here is the exact statement of this law:

Simple bodies A , B , and C may by uniting in various proportions form various compounds M , M' , . . . The masses of the bodies A , B , and C combining to form the compound M are to one another as the three numbers a , b , and c . Then the masses of the elements A , B , and C combining to form the compound M' will be to one another as the numbers xa , yb , and zc (x , y , and z being three whole numbers).

Is this law perhaps subject to experimental test? Chemical analysis will make us acquainted with the chemical composition of the body M' not exactly but with a certain approximation. The uncertainty of the results obtained can be extremely small; it will never be strictly zero. Now, in whatever relations the elements A , B , and C are combined within the compound M' , we can always represent these relations, with as close an approximation as you please, by the mutual relations of three products xa , yb , and zc , where x , y , and z are whole numbers; in other words, whatever the results given by the chemical analysis of the compound M' , we are always sure to find three integers x , y , and z thanks to which the law of multiple proportions will be verified with a precision greater than that of the experiment. Therefore, no chemical analysis, no matter how re-

finer, will ever be able to show the law of multiple proportions to be wrong.

In like manner, all crystallography rests entirely on the "law of rational indices" which is formulated in the following way:

A trihedral being formed by three faces of a crystal, a fourth face cuts the three edges of this trihedral at distances from the summit which are proportional to one another as three given numbers, the parameters of the crystal. Any other face whatsoever should cut these same edges at distances from the summit which are to one another as xa , yb , and zc , where x , y , and z are three integers, the indices of the new face of the crystal.

The most perfect protractor determines the direction of a crystal's face only with a certain degree of approximation; the relations among the three segments that such a face makes on the edges of the fundamental trihedral are always able to get by with a certain error; now, however small this error is, we can always choose three numbers x , y , and z such that the mutual relations of these segments are represented with the least amount of error by the mutual relations of the three numbers xa , yb , and zc ; the crystallographer who would claim that the law of rational indices is made justifiable by his protractor would surely not have understood the very meaning of the words he is employing.

The law of multiple proportions and the law of rational indices are mathematical statements deprived of all physical meaning. A mathematical statement has physical meaning only if it retains a meaning when we introduce the word "nearly" or "approximately." This is not the case with the statements we have just alluded to. Their object really is to assert that certain relations are *commensurable* numbers. They would degenerate into mere truisms if they were made to declare that these relations are approximately commensurable, for any incommensurable relation whatever is always approximately commensurable; it is even as near as you please to being commensurable.

Therefore, it would be absurd to wish to subject certain principles of mechanics to *direct* experimental test; it would be absurd to subject the law of multiple proportions or the law of rational indices to this *direct* test.

Does it follow that these hypotheses placed beyond the reach of direct experimental refutation have nothing more to fear from experiment? That they are guaranteed to remain immutable no matter what discoveries observation has in store for us? To pretend so would be a serious error.

Taken in isolation these different hypotheses have no experimental meaning; there can be no question of either confirming or contradicting them by experiment. But these hypotheses enter as essential foundations into the construction of certain theories of rational mechanics, of chemical theory, of crystallography. The object of these theories is to represent experimental laws; they are schematisms intended essentially to be compared with facts.

Now this comparison might some day very well show us that one of our representations is ill adjusted to the realities it should picture, that the

corrections which come and complicate our schematism do not produce sufficient concordance between this schematism and the facts, that the theory accepted for a long time without dispute should be rejected, and that an entirely different theory should be constructed on entirely different or new hypotheses. On that day some one of our hypotheses, which taken in isolation defied direct experimental refutation, will crumble with the system it supported under the weight of the contradictions inflicted by reality on the consequences of this system taken as a whole.¹¹

In truth, hypotheses which by themselves have no physical meaning undergo experimental testing in exactly the same manner as other hypotheses. Whatever the nature of the hypothesis is, we have seen at the beginning of this chapter that it is never in isolation contradicted by experiment; experimental contradiction always bears as a whole on the entire group constituting a theory without any possibility of designating which proposition in this group should be rejected.

There thus disappears what might have seemed paradoxical in the following assertion: Certain physical theories rest on hypotheses which do not by themselves have any physical meaning.

10 | Good Sense Is the Judge of Hypotheses Which Ought to Be Abandoned

When certain consequences of a theory are struck by experimental contradiction, we learn that this theory should be modified but we are not told by the experiment what must be changed. It leaves to the physicist the task of finding out the weak spot that impairs the whole system. No absolute principle directs this inquiry, which different physicists may conduct in very different ways without having the right to accuse one another of illogicality. For instance, one may be obliged to safeguard certain fundamental hypotheses while he tries to reestablish harmony between the consequences of the theory and the facts by complicating the schematism in which these hypotheses are applied, by invoking various causes of error, and by multiplying corrections. The next physicist, disdainful of these complicated artificial procedures, may decide to change some one of the essential assumptions supporting the entire system. The first physicist does not have the right to condemn in advance the boldness of the second one, nor does the latter have the right to treat the timidity of the first physicist as absurd. The methods they follow are justifiable only by experiment, and if they both succeed in satisfying the requirements of experiment each is logically permitted to declare himself content with the work that he has accomplished.

That does not mean that we cannot very properly prefer the work of one of the two to that of the other. Pure logic is not the only rule for our

judgments; certain opinions which do not fall under the hammer of the principle of contradiction are in any case perfectly unreasonable. These motives which do not proceed from logic and yet direct our choices, these "reasons which reason does not know" and which speak to the ample "mind of finesse" but not to the "geometric mind," constitute what is appropriately called good sense.

Now, it may be good sense that permits us to decide between two physicists. It may be that we do not approve of the haste with which the second one upsets the principles of a vast and harmoniously constructed theory whereas a modification of detail, a slight correction, would have sufficed to put these theories in accord with the facts. On the other hand, it may be that we may find it childish and unreasonable for the first physicist to maintain obstinately at any cost, at the price of continual repairs and many tangled-up stays, the worm-eaten columns of a building tottering in every part, when by razing these columns it would be possible to construct a simple, elegant, and solid system.

But these reasons of good sense do not impose themselves with the same implacable rigor that the prescriptions of logic do. There is something vague and uncertain about them; they do not reveal themselves at the same time with the same degree of clarity to all minds. Hence, the possibility of lengthy quarrels between the adherents of an old system and the partisans of a new doctrine, each camp claiming to have good sense on its side, each party finding the reasons of the adversary inadequate. The history of physics would furnish us with innumerable illustrations of these quarrels at all times and in all domains. Let us confine ourselves to the tenacity and ingenuity with which Biot by a continual bestowal of corrections and accessory hypotheses maintained the emissionist doctrine in optics, while Fresnel opposed this doctrine constantly with new experiments favoring the wave theory.

In any event this state of indecision does not last forever. The day arrives when good sense comes out so clearly in favor of one of the two sides that the other side gives up the struggle even though pure logic would not forbid its continuation. After Foucault's experiment had shown that light traveled faster in air than in water, Biot gave up supporting the emission hypothesis; strictly, pure logic would not have compelled him to give it up, for Foucault's experiment was *not* the crucial experiment that Arago thought he saw in it, but by resisting wave optics for a longer time Biot would have been lacking in good sense.

Since logic does not determine with strict precision the time when an inadequate hypothesis should give way to a more fruitful assumption, and since recognizing this moment belongs to good sense, physicists may hasten this judgment and increase the rapidity of scientific progress by trying consciously to make good sense within themselves more lucid and more vigilant. Now nothing contributes more to entangle good sense and to disturb its insight than passions and interests. Therefore, nothing will

delay the decision which should determine a fortunate reform in a physical theory more than the vanity which makes a physicist too indulgent towards his own system and too severe towards the system of another. We are thus led to the conclusion so clearly expressed by Claude Bernard: The sound experimental criticism of a hypothesis is subordinated to certain moral conditions; in order to estimate correctly the agreement of a physical theory with the facts, it is not enough to be a good mathematician and skillful experimenter; one must also be an impartial and faithful judge.

■ | Notes

1. Claude Bernard, *Introduction à [l'étude de] la Médecine expérimentale* (Paris, 1865), p. 63. (Translator's note: Translated into English by H. C. Greene, *An Introduction to the Study of Experimental Medicine* [New York: Henry Schuman, 1949].)
2. Claude Bernard, *Introduction à [l'étude de] la Médecine expérimentale* (Paris, 1865), p. 64.
3. *ibid.*, p. 70.
4. *ibid.*, p. 67.
5. G. Milhaud, "La Science rationnelle," *Revue de Métaphysique et de Morale*, IV (1896), 280. Reprinted in *Le Rationnel* (Paris, 1898), p. 45.
6. H. Poincaré, "Sur les Principes de la Mécanique," *Bibliothèque du Congrès international de Philosophie*, III: *Logique et Histoire des Sciences* (Paris, 1901), p. 457; "Sur la valeur objective des théories physiques," *Revue de Métaphysique et de Morale*, X (1902), 263; *La Science et l'Hypothèse*, p. 110.
7. E. Le Roy, "Un positivisme nouveau," *Revue de Métaphysique et de Morale*, IX (1901), 143-144.
8. H. Poincaré, "Sur les Principes de la Mécanique," *Bibliothèque du Congrès international de Philosophie*, Sec. III: "Logique et Histoire des Sciences" (Paris, (1901), pp. 475, 491.
9. *ibid.*, pp. 472ff.
10. P. Duhem, *Le Mixte et la Combinaison chimique: Essai sur l'évolution d'une idée* (Paris, 1902), pp. 159-161.
11. At the International Congress of Philosophy held in Paris in 1900, M. Poincaré developed this conclusion: "Thus is explained how experiment may have been able to edify (or suggest) the principles of mechanics, but will never be able to overthrow them." Against this conclusion, M. Hadamard offered various remarks, among them the following: "Moreover, in conformity with a remark of M. Duhem, it is not an isolated hypothesis but the whole group of the hypotheses of mechanics that we can try to verify experimentally." *Revue de Métaphysique et de Morale*, VIII (1900), 559.

Two Dogmas of Empiricism

Modern empiricism has been conditioned in large part by two dogmas. One is a belief in some fundamental cleavage between truths which are *analytic*, or grounded in meanings independently of matters of fact, and truths which are *synthetic*, or grounded in fact. The other dogma is *reductionism*: the belief that each meaningful statement is equivalent to some logical construct upon terms which refer to immediate experience. Both dogmas, I shall argue, are ill-founded. One effect of abandoning them is, as we shall see, a blurring of the supposed boundary between speculative metaphysics and natural science. Another effect is a shift toward pragmatism.

1 | Background for Analyticity

Kant's cleavage between analytic and synthetic truths was foreshadowed in Hume's distinction between relations of ideas and matters of fact, and in Leibniz's distinction between truths of reason and truths of fact. Leibniz spoke of the truths of reason as true in all possible worlds. Picturesqueness aside, this is to say that the truths of reason are those which could not possibly be false. In the same vein we hear analytic statements defined as statements whose denials are self-contradictory. But this definition has small explanatory value; for the notion of self-contradictoriness, in the quite broad sense needed for this definition of analyticity, stands in exactly the same need of clarification as does the notion of analyticity itself. The two notions are the two sides of a single dubious coin.

Kant conceived of an analytic statement as one that attributes to its subject no more than is already conceptually contained in the subject.

FROM W. V. Quine, *From a Logical Point of View* (Cambridge, Mass.: Harvard University Press, 1953), 20-46. Originally published in *Philosophical Review* 60 (1951): 20-45.

This formulation has two shortcomings: it limits itself to statements of subject-predicate form, and it appeals to a notion of containment which is left at a metaphorical level. But Kant's intent, evident more from the use he makes of the notion of analyticity than from his definition of it, can be restated thus: a statement is analytic when it is true by virtue of meanings and independently of fact. Pursuing this line, let us examine the concept of *meaning* which is presupposed.

Meaning, let us remember, is not to be identified with naming.¹ Frege's example of 'Evening Star' and 'Morning Star', and Russell's of 'Scott' and 'the author of *Waverley*', illustrate that terms can name the same thing but differ in meaning. The distinction between meaning and naming is no less important at the level of abstract terms. The terms '9' and 'the number of the planets' name one and the same abstract entity but presumably must be regarded as unlike in meaning; for astronomical observation was needed, and not mere reflection on meanings, to determine the sameness of the entity in question.

The above examples consist of singular terms, concrete and abstract. With general terms, or predicates, the situation is somewhat different but parallel. Whereas a singular term purports to name an entity, abstract or concrete, a general term does not; but a general term is *true* of an entity, or of each of many, or of none.² The *class* of all entities of which a general term is true is called the *extension* of the term. Now paralleling the contrast between the meaning of a singular term and the entity named, we must distinguish equally between the meaning of a general term and its extension. The general terms 'creature with a heart' and 'creature with kidneys', for example, are perhaps alike in extension but unlike in meaning.

Confusion of meaning with extension, in the case of general terms, is less common than confusion of meaning with naming in the case of singular terms. It is indeed a commonplace in philosophy to oppose intension (or meaning) to extension, or, in a variant vocabulary, connotation to denotation.

The Aristotelian notion of essence was the forerunner, no doubt, of the modern notion of intension or meaning. For Aristotle it was essential in men to be rational, accidental to be two-legged. But there is an important difference between this attitude and the doctrine of meaning. From the latter point of view it may indeed be conceded (if only for the sake of argument) that rationality is involved in the meaning of the word 'man' while two-leggedness is not; but two-leggedness may at the same time be viewed as involved in the meaning of 'biped' while rationality is not. Thus from the point of view of the doctrine of meaning it makes no sense to say of the actual individual, who is at once a man and a biped, that his rationality is essential and his two-leggedness accidental or vice versa. Things had essences, for Aristotle, but only linguistic forms have meanings. Meaning is what essence becomes when it is divorced from the object of reference and wedded to the word.

For the theory of meaning a conspicuous question is the nature of its objects: what sort of things are meanings? A felt need for meant entities may derive from an earlier failure to appreciate that meaning and reference are distinct. Once the theory of meaning is sharply separated from the theory of reference, it is a short step to recognizing as the primary business of the theory of meaning simply the synonymy of linguistic forms and the analyticity of statements; meanings themselves, as obscure intermediary entities, may well be abandoned.³

The problem of analyticity then confronts us anew. Statements which are analytic by general philosophical acclaim are not, indeed, far to seek. They fall into two classes. Those of the first class, which may be called *logically true*, are typified by:

- (1) No unmarried man is married.

The relevant feature of this example is that it not merely is true as it stands, but remains true under any and all reinterpretations of 'man' and 'married'. If we suppose a prior inventory of *logical* particles, comprising 'no', 'un-', 'not', 'if', 'then', 'and', etc., then in general a logical truth is a statement which is true and remains true under all reinterpretations of its components other than the logical particles.

But there is also a second class of analytic statements, typified by:

- (2) No bachelor is married.

The characteristic of such a statement is that it can be turned into a logical truth by putting synonyms for synonyms; thus (2) can be turned into (1) by putting 'unmarried man' for its synonym 'bachelor'. We still lack a proper characterization of this second class of analytic statements, and therewith of analyticity generally, inasmuch as we have had in the above description to lean on a notion of "synonymy" which is no less in need of clarification than analyticity itself.

In recent years Carnap has tended to explain analyticity by appeal to what he calls state-descriptions.⁴ A state-description is any exhaustive assignment of truth values to the atomic, or noncompound, statements of the language. All other statements of the language are, Carnap assumes, built up of their component clauses by means of the familiar logical devices, in such a way that the truth value of any complex statement is fixed for each state-description by specifiable logical laws. A statement is then explained as analytic when it comes out true under every state description. This account is an adaptation of Leibniz's "true in all possible worlds." But note that this version of analyticity serves its purpose only if the atomic statements of the language are, unlike 'John is a bachelor' and 'John is married', mutually independent. Otherwise there would be a state-description which assigned truth to 'John is a bachelor' and to 'John is married', and consequently 'No bachelors are married' would turn out synthetic rather than analytic under the proposed criterion. Thus the criterion of

analyticity in terms of state-descriptions serves only for languages devoid of extralogical synonym-pairs, such as 'bachelor' and 'unmarried man'—synonym-pairs of the type which give rise to the "second class" of analytic statements. The criterion in terms of state-descriptions is a reconstruction at best of logical truth, not of analyticity.

I do not mean to suggest that Carnap is under any illusions on this point. His simplified model language with its state-descriptions is aimed primarily not at the general problem of analyticity but at another purpose, the clarification of probability and induction. Our problem, however, is analyticity; and here the major difficulty lies not in the first class of analytic statements, the logical truths, but rather in the second class, which depends on the notion of synonymy.

2 | Definition

There are those who find it soothing to say that the analytic statements of the second class reduce to those of the first class, the logical truths, by definition; 'bachelor', for example, is *defined* as 'unmarried man'. But how do we find that 'bachelor' is defined as 'unmarried man'? Who defined it thus, and when? Are we to appeal to the nearest dictionary, and accept the lexicographer's formulation as law? Clearly this would be to put the cart before the horse. The lexicographer is an empirical scientist, whose business is the recording of antecedent facts; and if he glosses 'bachelor' as 'unmarried man' it is because of his belief that there is a relation of synonymy between those forms, implicit in general or preferred usage prior to his own work. The notion of synonymy presupposed here has still to be clarified, presumably in terms relating to linguistic behavior. Certainly the "definition" which is the lexicographer's report of an observed synonymy cannot be taken as the ground of the synonymy.

Definition is not, indeed, an activity exclusively of philologists. Philosophers and scientists frequently have occasion to "define" a recondite term by paraphrasing it into terms of a more familiar vocabulary. But ordinarily such a definition, like the philologist's, is pure lexicography, affirming a relation of synonymy antecedent to the exposition in hand.

Just what it means to affirm synonymy, just what the interconnections may be which are necessary and sufficient in order that two linguistic forms be properly describable as synonymous, is far from clear; but, whatever these interconnections may be, ordinarily they are grounded in usage. Definitions reporting selected instances of synonymy come then as reports upon usage.

There is also, however, a variant type of definitional activity which does not limit itself to the reporting of preëxisting synonymies. I have in mind what Carnap calls *explication*—an activity to which philosophers are given, and scientists also in their more philosophical moments. In explication the purpose is not merely to paraphrase the definiendum into an

outright synonym, but actually to improve upon the definiendum by refining or supplementing its meaning.* But even explication, though not merely reporting a preëxisting synonymy between definiendum and definiens, does rest nevertheless on other preëxisting synonymies. The matter may be viewed as follows. Any word worth explicating has some contexts which, as wholes, are clear and precise enough to be useful; and the purpose of explication is to preserve the usage of these favored contexts while sharpening the usage of other contexts. In order that a given definition be suitable for purposes of explication, therefore, what is required is not that the definiendum in its antecedent usage be synonymous with the definiens, but just that each of these favored contexts of the definiendum, taken as a whole in its antecedent usage, be synonymous with the corresponding context of the definiens.

Two alternative definientia may be equally appropriate for the purposes of a given task of explication and yet not be synonymous with each other; for they may serve interchangeably within the favored contexts but diverge elsewhere. By cleaving to one of these definientia rather than the other, a definition of explicative kind generates, by fiat, a relation of synonymy between definiendum and definiens which did not hold before. But such a definition still owes its explicative function, as seen, to preëxisting synonymies.

There does, however, remain still an extreme sort of definition which does not hark back to prior synonymies at all: namely, the explicitly conventional introduction of novel notations for purposes of sheer abbreviation. Here the definiendum becomes synonymous with the definiens simply because it has been created expressly for the purpose of being synonymous with the definiens. Here we have a really transparent case of synonymy created by definition; would that all species of synonymy were as intelligible. For the rest, definition rests on synonymy rather than explaining it.

The word 'definition' has come to have a dangerously reassuring sound, owing no doubt to its frequent occurrence in logical and mathematical writings. We shall do well to digress now into a brief appraisal of the role of definition in formal work.

In logical and mathematical systems either of two mutually antagonistic types of economy may be striven for, and each has its peculiar practical utility. On the one hand we may seek economy of practical

* The *definiendum* is the word or phrase to be defined; the *definiens* is the definition (that which does the defining). As Quine notes, philosophers and scientists often seek to explicate—to clarify and analyze—the notions they define. Accordingly, the definiens of an explication will not be exactly equivalent in meaning to the definiendum but will offer an improved, more precise version of it. Sometimes this involves analyzing the original definiendum into several distinct concepts. See, for example, the attempts to explicate the notions of confirmation, explanation, and reduction in chapters 5, 6, and 8 of this volume.

expression—ease and brevity in the statement of multifarious relations. This sort of economy calls usually for distinctive concise notations for a wealth of concepts. Second, however, and oppositely, we may seek economy in grammar and vocabulary; we may try to find a minimum of basic concepts such that, once a distinctive notation has been appropriated to each of them, it becomes possible to express any desired further concept by mere combination and iteration of our basic notations. This second sort of economy is impractical in one way, since a poverty in basic idioms tends to a necessary lengthening of discourse. But it is practical in another way: it greatly simplifies theoretical discourse *about* the language, through minimizing the terms and the forms of construction wherein the language consists.

Both sorts of economy, though *prima facie* incompatible, are valuable in their separate ways. The custom has consequently arisen of combining both sorts of economy by forging in effect two languages, the one a part of the other. The inclusive language, though redundant in grammar and vocabulary, is economical in message lengths, while the part, called primitive notation, is economical in grammar and vocabulary. Whole and part are correlated by rules of translation whereby each idiom not in primitive notation is equated to some complex built up of primitive notation. These rules of translation are the so-called *definitions* which appear in formalized systems. They are best viewed not as adjuncts to one language but as correlations between two languages, the one a part of the other.

But these correlations are not arbitrary. They are supposed to show how the primitive notations can accomplish all purposes, save brevity and convenience, of the redundant language. Hence the definiendum and its definiens may be expected, in each case, to be related in one or another of the three ways lately noted. The definiens may be a faithful paraphrase of the definiendum into the narrower notation, preserving a direct synonymy⁸ as of antecedent usage; or the definiens may, in the spirit of explication, improve upon the antecedent usage of the definiendum; or finally, the definiendum may be a newly created notation, newly endowed with meaning here and now.

In formal and informal work alike, thus, we find that definition—except in the extreme case of the explicitly conventional introduction of new notations—hinges on prior relations of synonymy. Recognizing then that the notion of definition does not hold the key to synonymy and analyticity, let us look further into synonymy and say no more of definition.

3 | Interchangeability

A natural suggestion, deserving close examination, is that the synonymy of two linguistic forms consists simply in their interchangeability in all

contexts without change of truth value—interchangeability, in Leibniz's phrase, *salva veritate*.⁶ Note that synonyms so conceived need not even be free from vagueness, as long as the vaguenesses match.

But it is not quite true that the synonyms 'bachelor' and 'unmarried man' are everywhere interchangeable *salva veritate*. Truths which become false under substitution of 'unmarried man' for 'bachelor' are easily constructed with the help of 'bachelor of arts' or 'bachelor's buttons'; also with the help of quotation, thus:

'Bachelor' has less than ten letters.

Such counterinstances can, however, perhaps be set aside by treating the phrases 'bachelor of arts' and 'bachelor's buttons' and the quotation 'bachelor' each as a single indivisible word and then stipulating that the interchangeability *salva veritate* which is to be the touchstone of synonymy is not supposed to apply to fragmentary occurrences inside of a word. This account of synonymy, supposing it acceptable on other counts, has indeed the drawback of appealing to a prior conception of "word" which can be counted on to present difficulties of formulation in its turn. Nevertheless some progress might be claimed in having reduced the problem of synonymy to a problem of wordhood. Let us pursue this line a bit, taking "word" for granted.

The question remains whether interchangeability *salva veritate* (apart from occurrences within words) is a strong enough condition for synonymy, or whether, on the contrary, some heteronymous expressions might be thus interchangeable. Now let us be clear that we are not concerned here with synonymy in the sense of complete identity in psychological associations or poetic quality; indeed no two expressions are synonymous in such a sense. We are concerned only with what may be called *cognitive* synonymy. Just what this is cannot be said without successfully finishing the present study; but we know something about it from the need which arose for it in connection with analyticity in §1. The sort of synonymy needed there was merely such that any analytic statement could be turned into a logical truth by putting synonyms for synonyms. Turning the tables and assuming analyticity, indeed, we could explain cognitive synonymy of terms as follows (keeping to the familiar example): to say that 'bachelor' and 'unmarried man' are cognitively synonymous is to say no more nor less than that the statement:

(3) All and only bachelors are unmarried men

is analytic.⁷

What we need is an account of cognitive synonymy not presupposing analyticity—if we are to explain analyticity conversely with help of cognitive synonymy as undertaken in §1. And indeed such an independent

account of cognitive synonymy is at present up for consideration, namely, interchangeability *salva veritate* everywhere except within words. The question before us, to resume the thread at last, is whether such interchangeability is a sufficient condition for cognitive synonymy. We can quickly assure ourselves that it is, by examples of the following sort. The statement:

- (4) Necessarily all and only bachelors are bachelors

is evidently true, even supposing 'necessarily' so narrowly construed as to be truly applicable only to analytic statements. Then, if 'bachelor' and 'unmarried man' are interchangeable *salva veritate*, the result:

- (5) Necessarily all and only bachelors are unmarried men

of putting 'unmarried man' for an occurrence of 'bachelor' in (4) must, like (4), be true. But to say that (5) is true is to say that (3) is analytic, and hence that 'bachelor' and 'unmarried man' are cognitively synonymous.

Let us see what there is about the above argument that gives it its air of hocus-pocus. The condition of interchangeability *salva veritate* varies in its force with variations in the richness of the language at hand. The above argument supposes we are working with a language rich enough to contain the adverb 'necessarily', this adverb being so construed as to yield truth when and only when applied to an analytic statement. But can we condone a language which contains such an adverb? Does the adverb really make sense? To suppose that it does is to suppose that we have already made satisfactory sense of 'analytic'. Then what are we so hard at work on right now?

Our argument is not flatly circular, but something like it. It has the form, figuratively speaking, of a closed curve in space.

Interchangeability *salva veritate* is meaningless until relativized to a language whose extent is specified in relevant respects. Suppose now we consider a language containing just the following materials. There is an indefinitely large stock of one-place predicates (for example, 'F' where 'Fx' means that x is a man) and many-place predicates (for example, 'G' where 'Gxy' means that x loves y), mostly having to do with extralogical subject matter. The rest of the language is logical. The atomic sentences consist each of a predicate followed by one or more variables 'x', 'y', etc.; and the complex sentences are built up of the atomic ones by truth functions ('not', 'and', 'or', etc.) and quantification.⁸ In effect such a language enjoys the benefits also of descriptions and indeed singular terms generally, these being contextually definable in known ways.⁹ Even abstract singular terms naming classes, classes of classes, etc., are contextually definable in case the assumed stock of predicates includes the two-place predicate of class

membership.¹⁰ Such a language can be adequate to classical mathematics and indeed to scientific discourse generally, except in so far as the latter involves debatable devices such as contrary-to-fact conditionals or modal adverbs like 'necessarily'.¹¹ Now a language of this type is extensional, in this sense: any two predicates which agree extensionally (that is, are true of the same objects) are interchangeable *salva veritate*.¹²

In an extensional language, therefore, interchangeability *salva veritate* is no assurance of cognitive synonymy of the desired type. That 'bachelor' and 'unmarried man' are interchangeable *salva veritate* in an extensional language assures us of no more than that (3) is true. There is no assurance here that the extensional agreement of 'bachelor' and 'unmarried man' rests on meaning rather than merely on accidental matters of fact, as does the extensional agreement of 'creature with a heart' and 'creature with kidneys'.

For most purposes extensional agreement is the nearest approximation to synonymy we need care about. But the fact remains that extensional agreement falls far short of cognitive synonymy of the type required for explaining analyticity in the manner of §1. The type of cognitive synonymy required there is such as to equate the synonymy of 'bachelor' and 'unmarried man' with the analyticity of (3), not merely with the truth of (3).

So we must recognize that interchangeability *salva veritate*, if construed in relation to an extensional language, is not a sufficient condition of cognitive synonymy in the sense needed for deriving analyticity in the manner of §1. If a language contains an intensional adverb 'necessarily' in the sense lately noted, or other particles to the same effect, then interchangeability *salva veritate* in such a language does afford a sufficient condition of cognitive synonymy; but such a language is intelligible only in so far as the notion of analyticity is already understood in advance.

The effort to explain cognitive synonymy first, for the sake of deriving analyticity from it afterward as in §1, is perhaps the wrong approach. Instead we might try explaining analyticity somehow without appeal to cognitive synonymy. Afterward we could doubtless derive cognitive synonymy from analyticity satisfactorily enough if desired. We have seen that cognitive synonymy of 'bachelor' and 'unmarried man' can be explained as analyticity of (3). The same explanation works for any pair of one-place predicates, of course, and it can be extended in obvious fashion to many-place predicates. Other syntactical categories can also be accommodated in fairly parallel fashion. Singular terms may be said to be cognitively synonymous when the statement of identity formed by putting '=' between them is analytic. Statements may be said simply to be cognitively synonymous when their biconditional (the result of joining them by 'if and only if') is analytic.¹³ If we care to lump all categories into a single formulation, at the expense of assuming again the notion of "word" which was appealed to early in this section, we can describe any two linguistic forms as cognitively synonymous when the two forms are interchangeable (apart from

occurrences within "words") *salva* (no longer *veritate* but) *analyticitate*. Certain technical questions arise, indeed, over cases of ambiguity or homonymy; let us not pause for them, however, for we are already digressing. Let us rather turn our backs on the problem of synonymy and address ourselves anew to that of analyticity.

4 | Semantical Rules

Analyticity at first seemed most naturally definable by appeal to a realm of meanings. On refinement, the appeal to meanings gave way to an appeal to synonymy or definition. But definition turned out to be a will-o'-the-wisp, and synonymy turned out to be best understood only by dint of a prior appeal to analyticity itself. So we are back at the problem of analyticity.

I do not know whether the statement 'Everything green is extended' is analytic. Now does my indecision over this example really betray an incomplete understanding, an incomplete grasp of the "meanings", of 'green' and 'extended'? I think not. The trouble is not with 'green' or 'extended', but with 'analytic'.

It is often hinted that the difficulty in separating analytic statements from synthetic ones in ordinary language is due to the vagueness of ordinary language and that the distinction is clear when we have a precise artificial language with explicit "semantical rules." This, however, as I shall now attempt to show, is a confusion.

The notion of analyticity about which we are worrying is a purported relation between statements and languages: a statement *S* is said to be *analytic* for a language *L*, and the problem is to make sense of this relation generally, that is, for variable '*S*' and '*L*'. The gravity of this problem is not perceptibly less for artificial languages than for natural ones. The problem of making sense of the idiom '*S* is analytic for *L*', with variable '*S*' and '*L*', retains its stubbornness even if we limit the range of the variable '*L*' to artificial languages. Let me now try to make this point evident.

For artificial languages and semantical rules we look naturally to the writings of Carnap. His semantical rules take various forms, and to make my point I shall have to distinguish certain of the forms. Let us suppose, to begin with, an artificial language L_0 whose semantical rules have the form explicitly of a specification, by recursion or otherwise, of all the analytic statements of L_0 . The rules tell us that such and such statements, and only those, are the analytic statements of L_0 . Now here the difficulty is simply that the rules contain the word 'analytic', which we do not understand! We understand what expressions the rules attribute analyticity to, but we do not understand what the rules attribute to those expressions. In short, before we can understand a rule which begins 'A statement *S* is

analytic for language L_0 if and only if . . .', we must understand the general relative term 'analytic for'; we must understand ' S is analytic for L ' where ' S ' and ' L ' are variables.

Alternatively we may, indeed, view the so-called rule as a conventional definition of a new simple symbol 'analytic-for- L_0 ', which might better be written untendentiously as ' K ' so as not to seem to throw light on the interesting word 'analytic'. Obviously any number of classes K , M , N , etc. of statements of L_0 can be specified for various purposes or for no purpose; what does it mean to say that K , as against M , N , etc., is the class of the "analytic" statements of L_0 ?

By saying what statements are analytic for L_0 we explain 'analytic-for- L_0 ' but not 'analytic', not 'analytic for'. We do not begin to explain the idiom ' S is analytic for L ' with variable ' S ' and ' L ', even if we are content to limit the range of ' L ' to the realm of artificial languages.

Actually we do know enough about the intended significance of 'analytic' to know that analytic statements are supposed to be true. Let us then turn to a second form of semantical rule, which says not that such and such statements are analytic but simply that such and such statements are included among the truths. Such a rule is not subject to the criticism of containing the un-understood word 'analytic'; and we may grant for the sake of argument that there is no difficulty over the broader term 'true'. A semantical rule of this second type, a rule of truth, is not supposed to specify all the truths of the language; it merely stipulates, recursively or otherwise, a certain multitude of statements which, along with others unspecified, are to count as true. Such a rule may be conceded to be quite clear. Derivatively, afterward, analyticity can be demarcated thus: a statement is analytic if it is (not merely true but) true according to the semantical rule.

Still there is really no progress. Instead of appealing to an unexplained word 'analytic', we are now appealing to an unexplained phrase 'semantical rule'. Not every true statement which says that the statements of some class are true can count as a semantical rule—otherwise *all* truths would be "analytic" in the sense of being true according to semantical rules. Semantical rules are distinguishable, apparently, only by the fact of appearing on a page under the heading 'Semantical Rules'; and this heading is itself then meaningless.

We can say indeed that a statement is *analytic-for- L_0* if and only if it is true according to such and such specifically appended "semantical rules," but then we find ourselves back at essentially the same case which was originally discussed: ' S is analytic-for- L_0 if and only if. . .'. Once we seek to explain ' S is analytic for L ' generally for variable ' L ' (even allowing limitation of ' L ' to artificial languages), the explanation 'true according to the semantical rules of L ' is unavailing; for the relative term 'semantical rule of' is as much in need of clarification, at least, as 'analytic for'.

It may be instructive to compare the notion of semantical rule with

that of postulate. Relative to a given set of postulates, it is easy to say what a postulate is: it is a member of the set. Relative to a given set of semantical rules, it is equally easy to say what a semantical rule is. But given simply a notation, mathematical or otherwise, and indeed as thoroughly understood a notation as you please in point of the translations or truth conditions of its statements, who can say which of its true statements rank as postulates? Obviously the question is meaningless—as meaningless as asking which points in Ohio are starting points. Any finite (or effectively specifiable infinite) selection of statements (preferably true ones, perhaps) is as much a set of postulates as any other. The word 'postulate' is significant only relative to an act of inquiry; we apply the word to a set of statements just in so far as we happen, for the year or the moment, to be thinking of those statements in relation to the statements which can be reached from them by some set of transformations to which we have seen fit to direct our attention. Now the notion of semantical rule is as sensible and meaningful as that of postulate, if conceived in a similarly relative spirit—relative, this time, to one or another particular enterprise of schooling unversant persons in sufficient conditions for truth of statements of some natural or artificial language *L*. But from this point of view no one signalization of a subclass of the truths of *L* is intrinsically more a semantical rule than another; and, if 'analytic' means 'true by semantical rules', no one truth of *L* is analytic to the exclusion of another.¹⁴

It might conceivably be protested that an artificial language *L* (unlike a natural one) is a language in the ordinary sense *plus* a set of explicit semantical rules—the whole constituting, let us say, an ordered pair; and that the semantical rules of *L* then are specifiable simply as the second component of the pair *L*. But, by the same token and more simply, we might construe an artificial language *L* outright as an ordered pair whose second component is the class of its analytic statements; and then the analytic statements of *L* become specifiable simply as the statements in the second component of *L*. Or better still, we might just stop tugging at our bootstraps altogether.

Not all the explanations of analyticity known to Carnap and his readers have been covered explicitly in the above considerations, but the extension to other forms is not hard to see. Just one additional factor should be mentioned which sometimes enters: sometimes the semantical rules are in effect rules of translation into ordinary language, in which case the analytic statements of the artificial language are in effect recognized as such from the analyticity of their specified translations in ordinary language. Here certainly there can be no thought of an illumination of the problem of analyticity from the side of the artificial language.

From the point of view of the problem of analyticity the notion of an artificial language with semantical rules is a *feu follet* [will-o'-the-wisp] *par excellence*. Semantical rules determining the analytic statements of an

artificial language are of interest only in so far as we already understand the notion of analyticity; they are of no help in gaining this understanding.

Appeal to hypothetical languages of an artificially simple kind could conceivably be useful in clarifying analyticity, if the mental or behavioral or cultural factors relevant to analyticity—whatever they may be—were somehow sketched into the simplified model. But a model which takes analyticity merely as an irreducible character is unlikely to throw light on the problem of explicating analyticity.

It is obvious that truth in general depends on both language and extralinguistic fact. The statement 'Brutus killed Caesar' would be false if the world had been different in certain ways, but it would also be false if the word 'killed' happened rather to have the sense of 'began'. Thus one is tempted to suppose in general that the truth of a statement is somehow analyzable into a linguistic component and a factual component. Given this supposition, it next seems reasonable that in some statements the factual component should be null; and these are the analytic statements. But, for all its *a priori* reasonableness, a boundary between analytic and synthetic statements simply has not been drawn. That there is such a distinction to be drawn at all is an unempirical dogma of empiricists, a metaphysical article of faith.

5 | The Verification Theory and Reductionism

In the course of these somber reflections we have taken a dim view first of the notion of meaning, then of the notion of cognitive synonymy, and finally of the notion of analyticity. But what, it may be asked, of the verification theory of meaning? This phrase has established itself so firmly as a catchword of empiricism that we should be very unscientific indeed not to look beneath it for a possible key to the problem of meaning and the associated problems.

The verification theory of meaning, which has been conspicuous in the literature from Peirce onward, is that the meaning of a statement is the method of empirically confirming or infirming [disconfirming] it. An analytic statement is that limiting case which is confirmed no matter what.

As urged in §1, we can as well pass over the question of meanings as entities and move straight to sameness of meaning, or synonymy. Then what the verification theory says is that statements are synonymous if and only if they are alike in point of method of empirical confirmation or infirmation.

This is an account of cognitive synonymy not of linguistic forms generally, but of statements.¹⁵ However, from the concept of synonymy of statements we could derive the concept of synonymy for other linguistic

forms, by considerations somewhat similar to those at the end of §3. Assuming the notion of "word," indeed, we could explain any two forms as synonymous when the putting of the one form for an occurrence of the other in any statement (apart from occurrences within "words") yields a synonymous statement. Finally, given the concept of synonymy thus for linguistic forms generally, we could define analyticity in terms of synonymy and logical truth as in §1. For that matter, we could define analyticity more simply in terms of just synonymy of statements together with logical truth; it is not necessary to appeal to synonymy of linguistic forms other than statements. For a statement may be described as analytic simply when it is synonymous with a logically true statement.

So, if the verification theory can be accepted as an adequate account of statement synonymy, the notion of analyticity is saved after all. However, let us reflect. Statement synonymy is said to be likeness of method of empirical confirmation or infirmation. Just what are these methods which are to be compared for likeness? What, in other words, is the nature of the relation between a statement and the experiences which contribute to or detract from its confirmation?

The most naïve view of the relation is that it is one of direct report. This is *radical reductionism*. Every meaningful statement is held to be translatable into a statement (true or false) about immediate experience. Radical reductionism, in one form or another, well antedates the verification theory of meaning explicitly so called. Thus Locke and Hume held that every idea must either originate directly in sense experience or else be compounded of ideas thus originating; . . . we might rephrase this doctrine in semantical jargon by saying that a term, to be significant at all, must be either a name of a sense datum or a compound of such names or an abbreviation of such a compound. So stated, the doctrine remains ambiguous as between sense data as sensory events and sense data as sensory qualities; and it remains vague as to the admissible ways of compounding. Moreover, the doctrine is unnecessarily and intolerably restrictive in the term-by-term critique which it imposes. More reasonably, and without yet exceeding the limits of what I have called radical reductionism, we may take full statements as our significant units—thus demanding that our statements as wholes be translatable into sense-datum language, but not that they be translatable term by term.

This emendation would unquestionably have been welcome to Locke and Hume . . . , but historically it had to await an important reorientation in semantics—the reorientation whereby the primary vehicle of meaning came to be seen no longer in the term but in the statement. This reorientation, seen in Bentham and Frege, underlies Russell's concept of incomplete symbols defined in use;¹⁶ also it is implicit in the verification theory of meaning, since the objects of verification are statements.

Radical reductionism, conceived now with statements as units, set

itself the task of specifying a sense-datum language and showing how to translate the rest of significant discourse, statement by statement, into it. Carnap embarked on this project in the *Aufbau*.*

The language which Carnap adopted as his starting point was not a sense-datum language in the narrowest conceivable sense, for it included also the notations of logic, up through higher set theory. In effect it included the whole language of pure mathematics. The ontology implicit in it (that is, the range of values of its variables) embraced not only sensory events but classes, classes of classes, and so on. Empiricists there are who would boggle at such prodigality. Carnap's starting point is very parsimonious, however, in its extralogical or sensory part. In a series of constructions in which he exploits the resources of modern logic with much ingenuity, Carnap succeeds in defining a wide array of important additional sensory concepts which, but for his constructions, one would not have dreamed were definable on so slender a basis. He was the first empiricist who, not content with asserting the reducibility of science to terms of immediate experience, took serious steps toward carrying out the reduction.

If Carnap's starting point is satisfactory, still his constructions were, as he himself stressed, only a fragment of the full program. The construction of even the simplest statements about the physical world was left in a sketchy state. Carnap's suggestions on this subject were, despite their sketchiness, very suggestive. He explained spatio-temporal point-instants as quadruples of real numbers and envisaged assignment of sense qualities to point-instants according to certain canons. Roughly summarized, the plan was that qualities should be assigned to point-instants in such a way as to achieve the laziest world compatible with our experience. The principle of least action was to be our guide in constructing a world from experience.

Carnap did not seem to recognize, however, that his treatment of physical objects fell short of reduction not merely through sketchiness, but in principle. Statements of the form 'Quality q is at point-instant $xy; z; t$ ' were, according to his canons, to be apportioned truth values in such a

* The reference is to Carnap's first major work, *Der Logische Aufbau der Welt* (1928), published two years after Carnap joined the Vienna Circle. It has been translated into English by Rolf A. George as *The Logical Structure of the World* (Berkeley, Calif.: University of California Press, 1969). Carnap's work was inspired by Wittgenstein's *Tractatus Logico-Philosophicus* (1921) and Russell's doctrine of logical atomism, especially Russell's injunction to philosophers that, wherever possible, they should replace inferences to unknown entities with logical constructions out of known entities. In the *Aufbau*, Carnap tried to reduce all the concepts used in science and everyday life to the sensory qualities given in immediate experience. As Quine notes, Carnap's project was never completed and, by 1936, Carnap had given up the reductionist thesis that statements about the physical world can be translated into equivalent statements about immediate experience.

way as to maximize and minimize certain over-all features, and with growth of experience the truth values were to be progressively revised in the same spirit. I think this is a good schematization (deliberately oversimplified, to be sure) of what science really does; but it provides no indication, not even the sketchiest, of how a statement of the form 'Quality q is at x,y,z ;' could ever be translated into Carnap's initial language of sense data and logic. The connective 'is at' remains an added undefined connective; the canons counsel us in its use but not in its elimination.

Carnap seems to have appreciated this point afterward; for in his later writings he abandoned all notion of the translatability of statements about the physical world into statements about immediate experience. Reductionism in its radical form has long since ceased to figure in Carnap's philosophy.

But the dogma of reductionism has, in a subtler and more tenuous form, continued to influence the thought of empiricists. The notion lingers that to each statement, or each synthetic statement, there is associated a unique range of possible sensory events such that the occurrence of any of them would add to the likelihood of truth of the statement, and that there is associated also another unique range of possible sensory events whose occurrence would detract from that likelihood. This notion is of course implicit in the verification theory of meaning.

The dogma of reductionism survives in the supposition that each statement, taken in isolation from its fellows, can admit of confirmation or infirmation at all. My countersuggestion, issuing essentially from Carnap's doctrine of the physical world in the *Aufbau*, is that our statements about the external world face the tribunal of sense experience not individually but only as a corporate body.¹⁷

The dogma of reductionism, even in its attenuated form, is intimately connected with the other dogma—that there is a cleavage between the analytic and the synthetic. We have found ourselves led, indeed, from the latter problem to the former through the verification theory of meaning. More directly, the one dogma clearly supports the other in this way: as long as it is taken to be significant in general to speak of the confirmation and infirmation of a statement, it seems significant to speak also of a limiting kind of statement which is vacuously confirmed, *ipso facto*, come what may; and such a statement is analytic.

The two dogmas are, indeed, at root identical. We lately reflected that in general the truth of statements does obviously depend both upon language and upon extralinguistic fact; and we noted that this obvious circumstance carries in its train, not logically but all too naturally, a feeling that the truth of a statement is somehow analyzable into a linguistic component and a factual component. The factual component must, if we are empiricists, boil down to a range of confirmatory experiences. In the extreme case where the linguistic component is all that matters, a true statement is analytic. But I hope we are now impressed with how stubbornly

the distinction between analytic and synthetic has resisted any straightforward drawing. I am impressed also, apart from prefabricated examples of black and white balls in an urn, with how baffling the problem has always been of arriving at any explicit theory of the empirical confirmation of a synthetic statement. My present suggestion is that it is nonsense, and the root of much nonsense, to speak of a linguistic component and a factual component in the truth of any individual statement. Taken collectively, science has its double dependence upon language and experience; but this duality is not significantly traceable into the statements of science taken one by one.

The idea of defining a symbol in use was, as remarked, an advance over the impossible term-by-term empiricism of Locke and Hume. The statement, rather than the term, came with Bentham to be recognized as the unit accountable to an empiricist critique. But what I am now urging is that even in taking the statement as unit we have drawn our grid too finely. The unit of empirical significance is the whole of science.

6 | Empiricism without the Dogmas

The totality of our so-called knowledge or beliefs, from the most casual matters of geography and history to the profoundest laws of atomic physics or even of pure mathematics and logic, is a man-made fabric which impinges on experience only along the edges. Or, to change the figure, total science is like a field of force whose boundary conditions are experience. A conflict with experience at the periphery occasions readjustments in the interior of the field. Truth values have to be redistributed over some of our statements. Reevaluation of some statements entails reevaluation of others, because of their logical interconnections—the logical laws being in turn simply certain further statements of the system, certain further elements of the field. Having reevaluated one statement we must reevaluate some others, which may be statements logically connected with the first or may be the statements of logical connections themselves. But the total field is so underdetermined by its boundary conditions, experience, that there is much latitude of choice as to what statements to reevaluate in the light of any single contrary experience. No particular experiences are linked with any particular statements in the interior of the field, except indirectly through considerations of equilibrium affecting the field as a whole.

If this view is right, it is misleading to speak of the empirical content of an individual statement—especially if it is a statement at all remote from the experiential periphery of the field. Furthermore it becomes folly to seek a boundary between synthetic statements, which hold contingently on experience, and analytic statements, which hold come what may. Any

statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system. Even a statement very close to the periphery can be held true in the face of recalcitrant experience by pleading hallucination or by amending certain statements of the kind called logical laws. Conversely, by the same token, no statement is immune to revision. Revision even of the logical law of the excluded middle has been proposed as a means of simplifying quantum mechanics;^{*} and what difference is there in principle between such a shift and the shift whereby Kepler superseded Ptolemy, or Einstein Newton, or Darwin Aristotle?

For vividness I have been speaking in terms of varying distances from a sensory periphery. Let me try now to clarify this notion without metaphor. Certain statements, though *about* physical objects and not sense experience, seem peculiarly germane to sense experience—and in a selective way: some statements to some experiences, others to others. Such statements, especially germane to particular experiences, I picture as near the periphery. But in this relation of “germaneness” I envisage nothing more than a loose association reflecting the relative likelihood, in practice, of our choosing one statement rather than another for revision in the event of recalcitrant experience. For example, we can imagine recalcitrant experiences to which we would surely be inclined to accommodate our system by reevaluating just the statement that there are brick houses on Elm Street, together with related statements on the same topic. We can imagine other recalcitrant experiences to which we would be inclined to accommodate our system by reevaluating just the statement that there are no centaurs, along with kindred statements. A recalcitrant experience can, I have urged, be accommodated by any of various alternative reevaluations in various alternative quarters of the total system; but, in the cases which we are now imagining, our natural tendency to disturb the total system as little as possible would lead us to focus our revisions upon these specific statements concerning brick houses or centaurs. These statements are felt, therefore, to have a sharper empirical reference than highly theoretical statements of physics or logic or ontology. The latter statements may be thought of as relatively centrally located within the total network, meaning merely that little preferential connection with any particular sense data obtrudes itself.

As an empiricist I continue to think of the conceptual scheme of science as a tool, ultimately, for predicting future experience in the light of past experience. Physical objects are conceptually imported into the situation as convenient intermediaries—not by definition in terms of ex-

* The law of excluded middle asserts that, for any statement *p*, the statement *p* or not-*p* is a necessary truth. For a discussion of arguments against the law and the alleged relevance of quantum mechanics to its revisability, see the discussion of logic (pp. 378–81) in the subsection, “Gillies’ Criticisms of Duhem” in the commentary on chapter 3.

perience, but simply as irreducible posits¹⁸ comparable, epistemologically, to the gods of Homer. For my part I do, *qua* lay physicist, believe in physical objects and not in Homer's gods; and I consider it a scientific error to believe otherwise. But in point of epistemological footing the physical objects and the gods differ only in degree and not in kind. Both sorts of entities enter our conception only as cultural posits. The myth of physical objects is epistemologically superior to most in that it has proved more efficacious than other myths as a device for working a manageable structure into the flux of experience.

Positing does not stop with macroscopic physical objects. Objects at the atomic level are posited to make the laws of macroscopic objects, and ultimately the laws of experience, simpler and more manageable; and we need not expect or demand full definition of atomic and subatomic entities in terms of macroscopic ones, any more than definition of macroscopic things in terms of sense data. Science is a continuation of common sense, and it continues the common-sense expedient of swelling ontology to simplify theory.

Physical objects, small and large, are not the only posits. Forces are another example; and indeed we are told nowadays that the boundary between energy and matter is obsolete. Moreover, the abstract entities which are the substance of mathematics—ultimately classes and classes of classes and so on up—are another posit in the same spirit. Epistemologically these are myths on the same footing with physical objects and gods, neither better nor worse except for differences in the degree to which they expedite our dealings with sense experiences.

The over-all algebra of rational and irrational numbers is underdetermined by the algebra of rational numbers, but is smoother and more convenient; and it includes the algebra of rational numbers as a jagged or gerrymandered part.¹⁹ Total science, mathematical and natural and human, is similarly but more extremely underdetermined by experience. The edge of the system must be kept squared with experience; the rest, with all its elaborate myths or fictions, has as its objective the simplicity of laws.

Ontological questions, under this view, are on a par with questions of natural science.²⁰ Consider the question whether to countenance classes as entities. This, as I have argued elsewhere,²¹ is the question whether to quantify with respect to variables which take classes as values. Now Carnap ([3], p. 32n) has maintained that this is a question not of matters of fact but of choosing a convenient language form, a convenient conceptual scheme or framework for science. With this I agree, but only on the proviso that the same be conceded regarding scientific hypotheses generally. Carnap ([3], p. 32n) has recognized that he is able to preserve a double standard for ontological questions and scientific hypotheses only by assuming an absolute distinction between the analytic and the synthetic; and I need not say again that this is a distinction which I reject.²²

The issue over there being classes seems more a question of conven-

ient conceptual scheme; the issue over there being centaurs, or brick houses on Elm Street, seems more a question of fact. But I have been urging that this difference is only one of degree, and that it turns upon our vaguely pragmatic inclination to adjust one strand of the fabric of science rather than another in accommodating some particular recalcitrant experience. Conservatism figures in such choices, and so does the quest for simplicity.

Carnap, Lewis, and others take a pragmatic stand on the question of choosing between language forms, scientific frameworks; but their pragmatism leaves off at the imagined boundary between the analytic and the synthetic. In repudiating such a boundary I espouse a more thorough pragmatism. Each man is given a scientific heritage plus a continuing barrage of sensory stimulation; and the considerations which guide him in warping his scientific heritage to fit his continuing sensory promptings are, where rational, pragmatic.

Notes

1. See p. 9 of "On What There Is," in Quine [2].
2. See p. 10 of "On What There Is," and pp. 107-15 of "Logic and the Reification of Universals," both in Quine [2].
3. See pp. 11ff. of "On What There Is" and pp. 48 of "Meaning in Linguistics," both in Quine [2].
4. Carnap [1], pp. 9ff.; Carnap [2], pp. 70ff.
5. According to an important variant sense of 'definition', the relation preserved may be the weaker relation of mere agreement in reference; see p. 132 of "Notes on the Theory of Reference," in Quine [2]. But definition in this sense is better ignored in the present connection, being irrelevant to the question of synonymy.
6. Cf. Lewis [1], p. 373.
7. This is cognitive synonymy in a primary, broad sense. Carnap ([1], pp. 56ff) and Lewis ([2], pp. 83ff) have suggested how, once this notion is at hand, a narrower sense of cognitive synonymy which is preferable for some purposes can in turn be derived. But this special ramification of concept-building lies aside from the present purposes and must not be confused with the broad sort of cognitive synonymy here concerned.
8. Pp. 81ff. of "New Foundations for Mathematical Logic," in Quine [2], contain a description of just such a language, except that there happens there to be just one predicate, the two-place predicate ' \in '.
9. See pp. 5-8 of "On What There Is," in Quine [2]; also pp. 85f. of "New Foundations for Mathematical Logic" and 166f. of "Meaning and Existential Inference," both in Quine [2].
10. See p. 87 of "New Foundations For Mathematical Logic," in Quine [2].

11. On such devices see also "Reference and Modality," in Quine [2].
12. This is the substance of Quine [1], p. 121.
13. The 'if and only if' itself is intended in the truth functional sense. See Carnap [1], p. 14.
14. The foregoing paragraph was not part of the present essay as originally published. It was prompted by Martin (see References).
15. The doctrine can indeed be formulated with terms rather than statements as the units. Thus Lewis describes the meaning of a term as "a criterion in mind, by reference to which one is able to apply or refuse to apply the expression in question in the case of presented, or imagined, things or situations" (Lewis [2], p. 133).—For an instructive account of the vicissitudes of the verification theory of meaning, centered however on the question of the meaningfulness rather than synonymy or analyticity, see Hempel (see References).
16. See p. 6 of "On What There Is," in Quine [2].
17. This doctrine was well argued by Duhem, pp. 302–28 (see References). Or see Lowinger, pp. 132–40 (see References).
18. Cf. pp. 17f. of "On What There Is," in Quine [2].
19. Cf. p. 18 of "On What There Is," in Quine [2].
20. "L'ontologie fait corps avec la science elle-même et ne peut en être séparée." [Ontology is an integral part of science and cannot be separated from it.] Meyerson, p. 439 (see References).
21. See pp. 12f. of "On What There Is," and pp. 102ff. of "Logic and the Reification of Universals," both in Quine [2].
22. For an effective expression of further misgivings over this distinction, see White (see References).

■ | References

- Carnap, Rudolf [1], *Meaning and Necessity* (Chicago: University of Chicago Press, 1947).
- Carnap, Rudolf [2], *Logical Foundations of Probability* (Chicago: University of Chicago Press, 1950).
- Carnap, Rudolf [3], "Empiricism, Semantics, and Ontology," *Revue internationale de philosophie* 4 (1950): 20–40. Reprinted in Linsky.
- Duhem, Pierre, *La Théorie physique: son objet et sa structure* (Paris, 1906).
- Hempel, C. G., "Problems and Changes in the Empiricist Criterion of Meaning," *Revue internationale de philosophie* 4 (1950): 41–63. Reprinted in Linsky.
- Lewis, C. I. [1], *A Survey of Symbolic Logic* (Berkeley, 1918).
- Lewis, C. I. [2], *An Analysis of Knowledge and Valuation* (LaSalle, Ill.: Open Court, 1946).

Linsky, Leonard (ed.), *Semantics and the Philosophy of Language* (Urbana: University of Illinois Press, 1952).

Lowinger, Armand, *The Methodology of Pierre Duhem* (New York: Columbia University Press, 1941).

Martin, R. M., "On 'Analytic,'" *Philosophical Studies* 5 (1952): 42-47.

Meyerson, Émile, *Identité et réalité* (Paris, 1908; 4th ed., 1932).

Quine, W. V. [1], *Mathematical Logic* (New York: Norton, 1940; Cambridge, Mass.: Harvard University Press, 1947; rev. ed., Cambridge, Mass.: Harvard University Press, 1951).

Quine, W. V. [2], *From a Logical Point of View* (Cambridge, Mass.: Harvard University Press, 1953).

White, Morton, "The Analytic and the Synthetic: An Untenable Dualism," in Sidney Hook (ed.), *John Dewey: Philosopher of Science and Freedom* (New York: Dial Press, 1950), 316-30. Reprinted in Linsky.

The Duhem Thesis and the Quine Thesis

In current writings on the philosophy of science, reference is often made to what is called 'the Duhem-Quine thesis'. Really, however, this is something of a misnomer; for, as we shall see, the Duhem thesis differs in many important respects from the Quine thesis. In this chapter I will expound the two theses in turn and explain how they differ. I will conclude the chapter by suggesting that the phrase 'the Duhem-Quine thesis' could be used to refer to a thesis which combines elements from both the Duhem thesis and the Quine thesis. . . .

1 | Preliminary Exposition of the Thesis. The Impossibility of a Crucial Experiment

Of Duhem's many significant contributions to the philosophy of science, perhaps the most important was his formulation of what I will call the *Duhem thesis*. With his usual clarity and incisiveness, Duhem states this thesis as a section heading thus:

An Experiment in Physics Can Never Condemn an Isolated Hypothesis but Only a Whole Theoretical Group (1904-5, p. 183 [260]).*

Later in this section he expounds the thesis as follows:

In sum, the physicist can never subject an isolated hypothesis to experimental test, but only a whole group of hypotheses; when the experiment is in dis-

FROM Donald Gillies, *Philosophy of Science in the Twentieth Century* (Oxford: Blackwell Publishers, 1993), 98-116.

* All page references in square brackets are to the excerpts from Duhem and Quine in this volume.

agreement with his predictions, what he learns is that at least one of the hypotheses constituting this group is unacceptable and ought to be modified; but the experiment does not designate which one should be changed. (p. 187 [263]).

In order to discuss the Duhem thesis, it will be useful to introduce the notion of an *observation statement*. . . . Let us take an observation statement to be a statement which can provisionally be agreed to be either true or false on the basis of observation and experiment.

According to the Duhem thesis, an isolated hypothesis in physics (h, say) can never be falsified by an observation statement, O. As a generalisation covering all the hypotheses of physics, this is somewhat doubtful. Physics does appear to contain some falsifiable hypotheses. Consider, for example, Kepler's first law that planets move in ellipses with the Sun at one focus. Suppose that we observe a large number of positions of a given planet and that these do not lie on an ellipse of the requisite kind. We have then surely falsified Kepler's first law. The schema of falsification can be written, where 'not-h' is short for 'It is not the case that h':

If h, then O, but not-O, therefore not-h. (1)

This uses a logical law called *modus tollens*.

However, the Duhem thesis does apply to some hypotheses. . . . Consider, for example, Newton's first law of motion (T_1 , say). . . . We cannot find an O such that schema (1) above holds when we substitute T_1 for h.*

Newton's full theory (T, say) consisted of three laws of motion (T_1 , T_2 , and T_3) and the law of gravity, T_4 . So T was a conjunction of these four laws ($T = T_1 \& T_2 \& T_3 \& T_4$). Even from T by itself, however, we cannot derive any observable consequences regarding the solar system. To do so, we need to add to T a number of auxiliary hypotheses: for example, that no other forces but gravitational ones act on the planets, that the interplanetary attractions are small compared with those between the Sun and the planets, that the mass of the Sun is very much greater than that

* In the book from which this reading is excerpted, Gillies refers to an argument of Poincaré to support the claim that Newton's first law of motion is not falsifiable. Newton's first law has the form of a conditional statement: if there is no external force acting on a body, then the velocity of the body will not change. Thus, to falsify the law requires a body that accelerates even though it is free from any net external force. Poincaré argues that if we find such an apparent counterexample, we can always deny that the body is genuinely free from a net external force by attributing the acceleration to forces exerted by as-yet-undetected invisible molecules. In this way, the law will be protected from refutation. See Henri Poincaré, *Science and Hypothesis*, trans. W. J. Greenstreet (New York: Dover, 1952), 95-96; originally published in French (1902).

of the planets, and so on. Let us call the conjunction of such auxiliary hypotheses which are appropriate in a given case A. We now have the schema:

$$\begin{array}{l} \text{If } T_1 \& T_2 \& T_3 \& T_4 \& A, \text{ then } O, \text{ but not-}O, \\ \text{therefore not-}(T_1 \& T_2 \& T_3 \& T_4 \& A). \end{array} \quad (2)$$

Moreover, from not- $(T_1 \& T_2 \& T_3 \& T_4 \& A)$ it follows that at least one of the set $(T_1, T_2, T_3, T_4, \text{ or } A)$ is false, but we cannot say which one.

As the history of science shows, it is often a very real problem in scientific research to decide which one of a group of hypotheses should be changed. Consider, for example, Adams and Leverrier's discovery of Neptune in 1846. From Newton's theory T together with auxiliary hypotheses, astronomers were able to calculate the theoretical orbit of Uranus (the most distant planet then known). This theoretical orbit did not agree with the observed orbit. This meant that either T or one of the auxiliary hypotheses was false. Adams and Leverrier conjectured that the auxiliary hypothesis concerning the number of planets was in error. They postulated a new planet Neptune beyond Uranus, and calculated the mass and position it would have to have to cause the observed perturbations in Uranus's orbit. Neptune was duly observed on 23 September 1846 only 52' away from the predicted position.¹

This part of the story is quite well known, but there were some subsequent events which are also relevant to the Duhem thesis. Another difficulty which occupied astronomers at the time concerned the anomalous motion of the perihelion of Mercury, which was found to advance slightly faster than it should do according to standard theory. Leverrier tried the same approach that had proved successful in the case of the Uranus anomaly. He postulated a planet Vulcan nearer to the Sun than Mercury, with a mass, orbit, and so forth which would explain the advance in Mercury's perihelion. However, no such planet could be found.

The discrepancy here is very small. Newcomb in 1898 gave its value as $41.24'' \pm 2.09''$ per century; that is, less than an eightieth part of a degree per century. However, this tiny anomaly was explained with great success by the general theory of relativity (T'), which Einstein proposed in 1915 as a replacement for Newton's theory, T. The value of the anomalous advance of the perihelion of Mercury which followed from the general theory of relativity was 42.89" per century—a figure well within the bounds set by Newcomb. We see that, although the Uranus anomaly and the Mercury anomaly were *prima facie* very similar, success was obtained in one case by altering an auxiliary hypothesis, in the other by altering the main theory.

In the next section, Duhem goes on to draw an important consequence from his thesis. This section is in fact headed 'A "Crucial Experiment" Is Impossible in Physics' (1904-5, p. 188 [264]). Duhem uses the

term *crucial experiment* in something like the sense given by Bacon in the *Novum Organum* to his 'fact of the cross'. He formulates this notion of crucial experiment as follows: 'Enumerate all the hypotheses that can be made to account for this group of phenomena; then, by experimental contradiction eliminate all except one: the latter will no longer be a hypothesis, but will become a certainty' (ibid.). However, there is an obvious objection to crucial experiments in this strong sense: namely, that we can never be sure that we have listed all the hypotheses capable of explaining a group of phenomena. Duhem makes this point as follows:

Experimental contradiction does not have the power to transform a physical hypothesis into an indisputable truth; in order to confer this power on it, it would be necessary to enumerate completely the various hypotheses which may cover a determinate group of phenomena; but the physicist is never sure he has exhausted all the imaginable assumptions. (p. 190 [266])

In view of this difficulty, it seems desirable to adopt a rather weaker sense of crucial experiment, which may be defined as follows. Suppose we have two competing theories T_1 and T_2 . An experiment (E , say) is crucial between T_1 and T_2 , if T_1 predicts that E will give the result O and T_2 predicts that E will give the result not- O . If we perform E , and O occurs, then T_2 is eliminated. If we perform E , and not- O occurs, then T_1 is eliminated. In any event, one of the two theories will be eliminated by E , which is thus crucial for deciding between them. It does not of course follow that the successful theory is necessarily true, because there may be some, as yet unthought of, theory, T_3 , which differs from T_1 and T_2 but explains the whole matter much more satisfactorily.

Duhem's point is that if T_1 and T_2 are such that his thesis applies to them, then we cannot derive O from T_1 but only from T_1 and A , where A is a conjunction of auxiliary assumptions. So, if not- O is the result of the experiment, this does not demonstrate beyond doubt that T_1 should be eliminated in favour of T_2 . It could be that one of the auxiliary hypotheses in A is at fault.

Duhem illustrates this by what is perhaps the most famous example of an alleged crucial experiment in the history of science: Foucault's experiment, which was designed to decide between the wave theory and the particle theory of light. The wave theory of light predicted that the velocity of light in water should be less than its velocity in air, whereas the particle theory predicted that the velocity of light in water should be greater than its velocity in air. Foucault devised a method for measuring the velocity of light in water, and found that it was actually less than the velocity of light in air. Here, then, we seem to have a crucial experiment which decides definitely in favour of the wave theory of light. Indeed, some of Foucault's contemporaries, notably Arago, did maintain that Foucault's experiment was a crucial experiment in just this sense.

Duhem pointed out, however, that to derive from the particle theory that the velocity of light in water is greater than its velocity in air, we need, not just the assumption that light consists of particles (the fundamental hypothesis of the particle theory), but many auxiliary assumptions as well. The particle theory could always be saved by altering some of these auxiliary assumptions. As Duhem puts it: 'For it is not between two hypotheses, the emission and wave hypotheses, that Foucault's experiment judges trenchantly; it decides rather between two sets of theories each of which has to be taken as a whole, i.e. between two entire systems, Newton's optics and Huygens' optics' (p. 189 [265]). So, according to Duhem, Foucault's experiment is not a crucial experiment in a strictly logical sense. Yet, as we shall see in the next section, there is another, weaker sense in which the experiment is crucial, even for Duhem.

2 | Duhem's Criticisms of Conventionalism. His Theory of Good Sense (*le bon sens*)

Duhem is sometimes classified as a conventionalist as regards his philosophy of science, but he is certainly not a conventionalist in the sense of Le Roy and Poincaré. Indeed, he devotes two sections of his *Aim and Structure of Physical Theory* to criticising these thinkers very clearly and explicitly. He formulates their conventionalist position as follows: 'Certain fundamental hypotheses of physical theory cannot be contradicted by any experiment, because they constitute in reality definitions, and because certain expressions in the physicist's usage take their meaning only through them' (p. 209 [271]).

Duhem objects strongly to Poincaré's claim that the principles of Newtonian mechanics will never be given up, because they are the simplest conventions available and cannot be contradicted by experiment. According to Duhem, the study of the history of science makes any such claim highly dubious:

The history of science should show that it would be very imprudent for us to say concerning a hypothesis commonly accepted today: 'We are certain that we shall never be led to abandon it because of a new experiment, no matter how precise it is.' Yet M. Poincaré does not hesitate to make this assertion concerning the principles of mechanics. (p. 212 [274]; I have here slightly altered the standard English translation in the interests of clarity.)

Poincaré's mistake, according to Duhem, was to take each principle of mechanics singly and in isolation. It is indeed true that when a principle of mechanics—for example, Newton's first law of motion—is taken in this fashion, it cannot be either confirmed or refuted by experience. However,

by adding other hypotheses to any such principle, we get a group of hypotheses which can be compared with experience. Moreover, if the group in question is contradicted by the results of experiment and observation, it is possible to change any of the hypotheses of the group. We cannot say with Poincaré that certain fundamental hypotheses, because they are appropriately simple conventions, are above question and can never be altered. This is how Duhem puts the matter:

It would be absurd to wish to subject certain principles of mechanics to *direct* experimental test: . . .

Does it follow that these hypotheses placed beyond the reach of direct experimental refutation have nothing more to fear from experiment? That they are guaranteed to remain immutable no matter what discoveries observation has in store for us? To pretend so would be a serious error.

Taken in isolation these different hypotheses have no experimental meaning; there can be no question of either confirming or contradicting them by experiment. But these hypotheses enter as essential foundations into the construction of certain theories of rational mechanics . . . these theories . . . are schematisms intended essentially to be compared with facts.

Now this comparison might some day very well show us that one of our representations is ill-adjusted to the realities it should picture, that the corrections which come and complicate our schematism do not produce sufficient concordance between this schematism and the facts, that the theory accepted for a long time without dispute should be rejected, and that an entirely different theory should be constructed on entirely different or new hypotheses. On that day some one of our hypotheses, which taken in isolation defied direct experimental refutation, will crumble with the system it supported under the weight of the contradictions inflicted by reality on the consequences of this system taken as a whole. (pp. 215–16 [276–77])

Thus Duhem's position seems to me more accurately described as *modified falsification*, rather than *conventionalism*. Duhem claims that some hypotheses of physics, when taken in isolation, can defy direct experimental refutation. He is thus not a strict falsificationist. On the other hand, he denies that such a hypothesis is immune from revision in the light of experimental evidence. A hypothesis of this kind may be tested indirectly if it forms part of a system of hypotheses which can be compared with experiment and observation. Further, such a hypothesis may on some occasion 'crumble with the system it supported under the weight of contradictions inflicted by reality'. Duhem does not deny that 'among the theoretical elements . . . there is always a certain number which the physicists of a certain epoch agree in accepting without test and which they regard as beyond dispute' (p. 211 [273]). However, he is very concerned

to warn scientists against adopting too dogmatic an attitude towards any of their assumptions. His point is that, in the face of recalcitrant experience, the best way forward may be to alter one of the most entrenched assumptions. As he says:

Indeed, we must really guard ourselves against believing forever warranted those hypotheses which have become universally adopted conventions, and whose certainty seems to break through experimental contradiction by throwing the latter back on more doubtful assumptions. The history of physics shows us that very often the human mind has been led to overthrow such principles completely, though they have been regarded by common consent for centuries as inviolable axioms, and to rebuild its physical theories on new hypotheses. (p. 212 [273])

Duhem gives as an example the principle that light travels in a straight line. This was accepted as correct for hundreds—indeed, thousands—of years, but was eventually modified to explain certain diffraction effects.

Duhem even cites Newton's law of gravity as a law which is only provisional and may be changed in future. Unfortunately this passage has been accidentally omitted from the English edition of the *Aim and Structure of Physical Theory*. It is here translated from the French edition:

Of all the laws of physics, the one best verified by its innumerable consequences is surely the law of universal gravity; the most precise observations on the movements of the stars have not been able up to now to show it to be faulty. Is it, for all that, a definitive law? It is not, but a provisional law which has to be modified and completed unceasingly to make it accord with experience. (p. 267)

The episode of the anomalous motion of the perihelion of Mercury fits Duhem's analysis perfectly. It would surely have seemed reasonable to explain such a small discrepancy between Newton's theory and observation by altering some auxiliary assumption. In fact, however, the anomaly was only explained satisfactorily when Newton's whole theory of gravity was replaced by Einstein's general theory of relativity. Indeed, from a logical point of view, Duhem's philosophy of science can be seen as offering support to the Einsteinian revolution in physics. It therefore comes as a surprise to discover that Duhem rejected Einstein's theory of relativity in the most violent terms. In his 1915 booklet *La Science allemande* ('German Science'), Duhem argues that Einstein's theory of relativity must be considered as an aberration due to the lack of sound judgement of the German mind and its disrespect for reality. Admittedly, this booklet was written at a time when bitter nationalistic feelings were being generated by the First World War. Indeed, it belongs to a genre known as 'war literature', and is actually a relatively mild example of this unfortunate

species of writing. All the same, it is clear that Duhem did reject Einstein's theory of relativity in no uncertain terms.

So, as already observed, we find in both Duhem and Poincaré a contradiction between their philosophical views and their scientific practice. Duhem was led by philosophical considerations to the conclusion that Newtonian mechanics is provisional and may be altered in future; yet he repudiated the new Einsteinian mechanics.² Conversely, Poincaré suggested in his philosophical writings of 1902 that the principles of Newtonian mechanics were conventions so simple that they would never be given up; yet, only two years later, in 1904, he decided that Newtonian mechanics needed to be changed, and started work on the development of a new mechanics. Some light is thrown on these strange contradictions by one further element in the Duhem thesis which we have still to discuss. This is Duhem's theory of good sense (*le bon sens*).

Let us take the typical situation envisioned by the Duhem thesis. From a group of hypotheses, $\{h_1 \dots h_n\}$, say, a scientist has deduced O . Experiment or observation then shows that O is false. It follows that at least one of $\{h_1 \dots h_n\}$ is false. But which one or ones are false? Which hypothesis or hypotheses should the scientist try to change in order to re-establish the agreement between theory and experience? Duhem states quite categorically that logic by itself cannot help the scientist. As far as pure logic is concerned, the choice between the various hypotheses is entirely open. The scientist in reaching his decision must be guided by what Duhem calls 'good sense' (*le bon sens*):

Pure logic is not the only rule for our judgements; certain opinions which do not fall under the hammer of the principle of contradiction are in any case perfectly unreasonable. These motives which do not proceed from logic and yet direct our choices, these 'reasons which reason does not know' and which speak to the ample 'mind of finesse' but not to the 'geometric mind,' constitute what is appropriately called good sense. (1904-5, p. 217 [277-78])

Duhem imagines two scientists who, when faced with the experimental contradiction of a group of hypotheses, adopt different strategies. Scientist A alters a fundamental theory in the group, whereas scientist B alters some of the auxiliary assumptions. Both strategies are logically possible, and only good sense can enable us to decide between the two scientists. Thus, in the dispute between the particle theory of light and the wave theory of light, Biot, by a continual alteration and addition of auxiliary assumptions, tenaciously and ingeniously defended the particle theory, whereas Fresnel constantly devised new experiments favouring the wave theory. In the end, however, the dispute was resolved.

After Foucault's experiment had shown that light travelled faster in air than in water, Biot gave up supporting the emission hypothesis; strictly, pure logic

would not have compelled him to give it up, for Foucault's experiment was not the crucial experiment that Arago thought he saw in it. but by resisting wave optics for a longer time Biot would have been lacking in good sense. (p. 218, [278])

This passage in effect qualifies some of Duhem's earlier remarks about crucial experiments. Let us take two theories, T_1 and T_2 , which are both subject to the Duhem thesis; that is, which cannot be tested in isolation but only by adjoining further assumptions. In a strictly logical sense, there cannot be a crucial experiment which decides between T_1 and T_2 . The good sense of the scientific community can, however, lead it to judge that a particular experiment, such as Foucault's experiment, is in practice crucial in deciding the scientific controversy in favour of one of the two contending theories.

In his 1991 book (particularly chapters 4–6), Martin argues that 'life-long meditation on certain texts of Pascal shaped many of the most important and difficult features of Duhem's thought' (p. 101). In particular, Duhem's theory of good sense (*le bon sens*) was derived in part from Pascal. Indeed, in the passage introducing *le bon sens*, Duhem quotes part of Pascal's famous saying that the heart has its reasons which reason knows nothing of.³

Although Duhem was undoubtedly influenced by Pascal, it is possible to suggest factors of a more personal and psychological nature which may have led him to his theory of scientific good sense. As his writings on philosophy of science show, Duhem was a man of outstanding logical ability; yet, as a physicist, he was a failure. In almost every scientific controversy in which he was involved, he chose the wrong side, rejecting those theories such as atomism, Maxwell's electrodynamics, and Einstein's theory of relativity which were to prove successful and lead to scientific progress. Although Duhem stubbornly defended his erroneous scientific opinions, he must have known in his heart of hearts that he was not proving to be a successful scientist. Yet he must also have been aware of his own exceptional logical powers. This situation could only be explained by supposing that something in addition to pure logic was needed in order to become a successful scientist. Here, then, we have a possible psychological origin of Duhem's theory of scientific good sense: namely, that Duhem saw that good sense is necessary for a scientist precisely because he himself was lacking in good sense. Duhem's rejection of a new theory which agreed so well with his own philosophy of science (that is, Einstein's theory of relativity) is just another instance of that lack of good sense which unfortunately characterized Duhem's scientific career.

Poincaré, by contrast, was one of the great physicists of his generation, and was amply endowed with the scientific good sense which Duhem lacked. The contrast between the two men is particularly evident in their respective discussions of electrodynamics. As we have already remarked,

Duhem attacked Maxwell's theory harshly, and advocated the ideas of Helmholtz. Poincaré devotes a chapter (the thirteenth) of his 1902 book to electrodynamics. He begins (pp. 225–38) by discussing the theories of Ampère and Helmholtz and by mentioning the difficulties which he finds in these theories. Then, on p. 239, he introduces Maxwell's theory with the words: 'Such were the difficulties raised by the current theories, when Maxwell with a stroke of the pen caused them to vanish.' Subsequent developments completely endorsed Poincaré's support for Maxwell, while Helmholtz's ideas on electrodynamics, so strenuously advocated by Duhem, are now remembered only by a few erudite historians of science. It was Poincaré's scientific good sense which led him, contrary to the principles of his own conventionalist philosophy of science of 1902, to a modification of Newtonian mechanics.

Duhem's theory of good sense seems to me correct, but, at the same time, more in the nature of a problem, or a starting-point for further analysis, than of a final solution to the difficulty with which it deals. What factors contribute to forming scientific good sense? Why are some highly intelligent individuals like Duhem lacking in good sense? These are important questions. . . . In the next section, however, I will turn to a consideration of the Quine thesis.

3 | The Quine Thesis

In his famous 1951 article, 'Two Dogmas of Empiricism', Quine puts forward, with a reference to Duhem, a thesis which is related to Duhem's. Nonetheless, it seems to me that Quine's thesis is sufficiently different from Duhem's to make the conflation of the two intellectually unsatisfactory.⁴ I will next briefly describe the Quine thesis,⁵ and explain how it differs from the Duhem thesis.

The first obvious difference between Quine and Duhem is that Quine develops his views in the context of a discussion about whether a distinction can be drawn between analytic and synthetic statements, whereas Duhem does not even mention (let alone discuss) the analytic/synthetic problem.

[There are] two ways of defining an analytic statement. The first [is] due to Kant, who actually introduced the analytic/synthetic distinction. According to Kant, a statement is analytic if its predicate is contained in its subject. This formulation presupposes an Aristotelian analysis of statements into subject and predicate. It is not surprising that Frege, who rejected Aristotelian logic and introduced modern logic, should have proposed a new way of defining an analytic statement. Frege defines an analytic statement as one which is reducible to a truth of logic by means of explicit definitions. These two ways of defining an analytic statement

are both illustrated by the standard example of an analytic statement, namely 'All bachelors are unmarried'. But Quine defines analytic statement in yet a third way. He writes critically of 'a belief in some fundamental cleavage between truths which are *analytic*, or grounded in meanings independently of matters of fact, and truths which are *synthetic*, or grounded in fact' (1951, p. 20 [280]). In effect, Quine is here taking a sentence to be analytic if it is true in virtue of the meanings of the words it contains. This is the definition of 'analytic' which is adopted by most modern philosophers interested in the question. Once again it is admirably illustrated by the standard example: $S =$ 'All bachelors are unmarried'. Someone who knows the meanings of 'all', 'bachelors', 'are', and 'unmarried' will at once recognize that S is true, without having to make any empirical investigations into matters of fact. Thus S is analytic.

All this seems very convincing; yet Quine denies that the distinction between analytic and synthetic is a valid one. He writes:

It is obvious that truth in general depends on both language and extralinguistic fact. The statement 'Brutus killed Caesar' would be false if the world had been different in certain ways, but it would also be false if the word 'killed' happened rather to have the sense of 'begat'. Thus one is tempted to suppose in general that the truth of a statement is somehow analyzable into a linguistic component and a factual component. Given this supposition, it next seems reasonable that in some statements the factual component should be null; and these are the analytic statements. But, for all its *a priori* reasonableness, a boundary between analytic and synthetic statements simply has not been drawn. That there is such a distinction to be drawn at all is an unempirical dogma of empiricists, a metaphysical article of faith. (1951, pp. 36-7 [292])

The empiricists to whom Quine refers are, of course, the empiricists of the Vienna Circle, especially Carnap. . . . Their particular brand of empiricism (logical empiricism) did indeed involve drawing a distinction between analytic and synthetic statements. However, support for the distinction is not confined to some members of the empiricist camp. Kantians too support the distinction, which was indeed introduced by Kant himself.

But what has all this to do with the issues involving Duhem and conventionalism, which we have been discussing? We can begin to build a bridge by observing that the meanings given to sounds and inscriptions are determined purely by social convention. Indeed, the social conventions differ from one language to another. So if a sentence is true in virtue of the meanings of the words it contains (that is, is analytic), it is *a fortiori* true by convention. Thus if a law is analytic, it is true by convention. The converse may not hold, since it is conceivable that a law might be rendered true by a set of conventions which include not just linguistic con-

ventions concerning the meanings of words but also perhaps conventions connected with measuring procedures.

Duhem used his thesis against the claim that a particular scientific law was true by convention. It is now obvious that exactly the same argument could be used against the claim that the law is analytic. Indeed, Quine does argue against the analytic/synthetic distinction along just these lines.⁶

But to carry his argument through, Quine makes a claim (the Quine thesis) which is much stronger than the Duhem thesis. The key difference between the two theses is clearly expressed by Vuillemin as follows: 'Duhem's thesis ("D-thesis") has a limited and special scope not covering the field of physiology, for Claude Bernard's experiments are explicitly acknowledged as crucial. Quine's thesis ("Q-thesis") embraces the whole body of our knowledge' (1979, p. 599).

Duhem does indeed place explicit limitations on the scope of his thesis. He writes: 'The Experimental Testing of a Theory Does Not Have the Same Logical Simplicity in Physics as in Physiology' (1904-5, p. 180 [257]). He thinks that his thesis does not apply in physiology or in certain branches of chemistry, and defends it only for the hypotheses of physics. My own view is that Duhem is correct to limit the scope of his thesis, but wrong to identify its scope with that of a particular branch of science—namely, physics. There are in physics falsifiable laws—for example, Snell's law of refraction applied to glass—whereas physiology and chemistry no doubt contain hypotheses subject to the Duhem thesis. . . . For the moment, however, it is not of great importance where exact boundaries are drawn. The crucial point is that Duhem wanted to apply his thesis to some statements and not to others, whereas the Quine thesis is supposed to apply to any statement whatever.

This is closely connected with a second difference between the Duhem thesis and the Quine thesis. Duhem maintains that hypotheses in physics cannot be tested in isolation, but only as part of a group. However, his discussion makes clear that he places limits on the size of this 'group'. Quine, however, thinks that the group extends and ramifies until it includes the whole of human knowledge. Quine writes: 'The unit of empirical significance is the whole of science' (1951, p. 42 [296]); and again:

The totality of our so-called knowledge or beliefs, from the most casual matters of geography and history to the profoundest laws of atomic physics or even of pure mathematics and logic, is a man-made fabric which impinges on experience only along the edges. Or, to change the figure, total science is like a field of force whose boundary conditions are experience. A conflict with experience at the periphery occasions readjustments in the interior of the field. . . . But the total field is so underdetermined by its boundary conditions, experience, that there is much latitude of choice as to what statements to reevaluate in the light of any single contrary experience. No particular

experiences are linked with any particular statements in the interior of the field, except indirectly through considerations of equilibrium affecting the field as a whole. (pp. 42–3 [296])

The Quine thesis is stronger than the Duhem thesis, and, in my view, less plausible. Let us take, as a concrete example, one of the cases analysed earlier. Newton's first law cannot, taken in isolation, be compared with experience. Adams and Leverrier, however, used this law as one of a group of hypotheses from which they deduced conclusions about the orbit of Uranus. These conclusions disagreed with observation. Now the group of hypotheses used by Adams and Leverrier was, no doubt, fairly extensive, but it did not include the whole of science. Adams and Leverrier did not, for example, mention the assumption that bees collect nectar from flowers in order to make honey, although such an assumption might well have appeared in a contemporary scientific treatise dealing with a question in biology. We agree, then, with Quine that a single statement may not always be (to use his terminology) a 'unit of empirical significance'. But this does not mean that 'The unit of empirical significance is the whole of science' (1951, p. 42 [296]). A group of statements which falls considerably short of the whole of science may sometimes be a perfectly valid unit of empirical significance.

Another difference between Duhem and Quine is that Quine does not have a theory of scientific good sense. Let us take, for example, Quine's statement: 'Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system' (p. 43 [296–97]). It is easy to imagine how Duhem would have reacted to such an assertion when applied to a statement falling under his thesis. Duhem would have agreed that, from the point of view of *pure logic*, one can indeed hold a particular statement—for example, Newton's particle theory of light—to be true, come what may. However, someone who did so in certain evidential situations would be *lacking in good sense*, and indeed *perfectly unreasonable*.

Because Quine does not have a theory of good sense, he cannot give the Duhemian analysis which we have just sketched. Indeed, it is significant that his 1951 article, 'Two Dogmas of Empiricism', is reprinted in a collection entitled *From a Logical Point of View*. Where Quine does go beyond logic, it is towards pragmatism, though Quine's pragmatism is usually mentioned only in passing, rather than elaborated, as in the following passage: 'Each man is given a scientific heritage plus a continuing barrage of sensory stimulation, and the considerations which guide him in warping his scientific heritage to fit his continuing sensory promptings are, where rational, pragmatic' (p. 46 [299]).

Although the Duhem thesis is quite clearly distinct from the Quine thesis, it might still be possible—indeed, useful—to form a composite thesis containing some, but not all, elements from each of the two theses. The phrase *Duhem-Quine thesis* could then be validly used to denote this

composite thesis. In the last section of this chapter, I will elaborate a suggestion along these lines.

4 | The Duhem-Quine Thesis

Let us say that the *holistic thesis* applies to a particular hypothesis if that hypothesis cannot be refuted by observation and experiment when taken in isolation, but only when it forms part of a theoretical group. The differences between the Duhem and Quine theses concern the range of hypotheses to which the holistic thesis is applied and the extent of the 'theoretical group' for a hypothesis to which the holistic thesis does apply. In discussing these differences, I have so far sided with Duhem against Quine. There is one point, however, on which I would like to defend Quine against Duhem. Quine, as we have seen, extends the holistic thesis to mathematics and logic. Duhem, however, thought that mathematics and logic had a character quite different from that of physics. Crowe (1990) gives an excellent general account and critique of Duhem's views on the history and philosophy of mathematics. I will here confine myself to a brief account of some views concerning geometry and logic which Duhem expounded in his late work *La Science allemande* ('German Science').

Duhem begins his treatment of geometry with the following remarks:

Among the sciences of reasoning, arithmetic and geometry are the most simple and, consequently, the most completely finished; . . .

What is the source of their axioms? They are taken, it is usually said, from common sense knowledge (*connaissance commune*): that is to say that any man sane of mind is sure of their truth before having studied the science of which they will be the foundations. (1915, pp. 4–5)

Duhem agrees with this point of view. In fact, he holds what in 1915 was a very old-fashioned opinion, that the axioms of Euclid are established as true by common-sense knowledge (*connaissance commune*) or common sense (*le sens commun*) or intuitive knowledge (*connaissance intuitive*). A proposition from which Euclid's fifth postulate can be deduced is that, given a geometrical figure (say a triangle), there exists another geometrical figure similar to it but of a different size. Duhem argues that the intuitions of palaeolithic hunters of reindeer were sufficient to establish the truth of this proposition. As he says:

One can represent a plane figure by drawing, or a solid figure by sculpture, and the image can resemble the model perfectly, even though they have different sizes. This is a truth which was in no way doubted, in palaeolithic

times, by the hunters of reindeer on the banks of the Vézère. Now that figures can be similar without being equal, implies, as the geometric spirit demonstrates, the exact truth of Euclid's postulate. (pp. 115-16)

Naturally enough, this attitude to the foundations of geometry leads Duhem to criticize non-Euclidean geometry, and, in particular, Riemannian geometry. This is what he says:

Riemann's doctrine is a *rigorous algebra*, for all the theorems which it formulates are very precisely deduced from its basic postulates; so it satisfies the geometric spirit. It is not a *true geometry*, for, in putting forward its postulates, it is not concerned that their corollaries should agree at every point with the judgements, drawn from experience, which constitute our intuitive knowledge of space: it is therefore repugnant to common sense. (p. 118)

It is perhaps no accident that the non-Euclidean geometer cited by Duhem (namely, Riemann) was a German; for, as already remarked, *La Science allemande*, written in 1915, was an example of the war literature of the time, designed to denigrate the enemy nationality. Duhem attacks German scientists by claiming that, while they possess the geometric spirit (*l'esprit géométrique*), their theories contradict common sense (*le sens commun*) or *l'esprit de finesse*, which is Duhem's new term for something like his old notion of good sense.

Given this general point of view, it is not surprising that we find Duhem condemning the theory of relativity. He speaks of 'the principle of relativity such as has been conceived by an Einstein, a Max Abraham, a Minkowski, a Laue' (p. 135). Forgetting the contributions of his own compatriot Poincaré, he denounces relativity as a typical aberration of the German mind. As he says:

The fact that the principle of relativity confounds all the intuitions of common sense, does not arouse against it the mistrust of the German physicists—quite the contrary! To accept it is, by that very fact, to overturn all the doctrines where space, time, movement were treated, all the theories of mechanics and physics; such a devastation has nothing about it which can displease German thought; on the ground which it will have cleared of the ancient doctrines, the geometric spirit of the Germans will devote itself with a happy heart to rebuilding a whole new physics of which the principle of relativity will be the foundation. If this new physics, disdainful of common sense, runs counter to all that observation and experience have allowed to be constructed in the domain of celestial and terrestrial mechanics, the purely deductive method will only be more proud of the inflexible rigour with which it will have followed to the end the ruinous consequences of its postulate. (p. 136)

The development and acceptance of non-Euclidean geometry and relativity have rendered Duhem's attempt to found geometry on common sense untenable. It is surely now more reasonable to extend the holistic thesis from physics to geometry and to say that, in the face of recalcitrant observations, we have the option of altering postulates of geometry as well as postulates of physics. This is, after all, precisely what Einstein did when he devised his general theory of relativity.

The picture is the same when we turn from geometry to logic. . . . Duhem [claimed] that 'There is a general method of deduction; Aristotle has formulated its laws for all time (*pour toujours*)' (p. 58). Yet by 1915 the new logic of Frege, Peano, and Russell had clearly superseded Aristotelian logic. Moreover, Brouwer had criticized some of the standard logical laws, and suggested his alternative intuitionistic approach. Quine writes: 'Revision even of the logical law of the excluded middle has been proposed as a means of simplifying quantum mechanics' (1951, p. 43 [297]). Admittedly the new 'quantum logic' has not proved very successful in resolving the paradoxes of microphysics; but there is no reason in principle why a change of this kind should not prove efficacious in some scientific context. In artificial intelligence, non-standard logics (for example, non-monotonic logics) are being devised in order to model particular forms of intelligent reasoning, and this programme has met with some success. Thus it seems reasonable to extend the holistic thesis to include logic as well and to allow the possibility of altering logical laws as well as scientific laws to explain recalcitrant observations.

I am now in a position to formulate what I will call the *Duhem-Quine thesis*, which combines what seem to me the best aspects of the Duhem thesis and the Quine thesis. It will be convenient to divide the statement in two parts.

- A The holistic thesis applies to any high-level (level 2) theoretical hypotheses, whether of physics or of other sciences, or even of mathematics and logic. (A incorporates ideas from the Quine thesis.)
- B The group of hypotheses under test in any given situation is in practice limited, and does not extend to the whole of human knowledge. Quine's claim that 'Any statement can be held to be true come what may, if we make drastic enough adjustments elsewhere in the system' (1951, p. 43 [296-97]) is true from a purely logical point of view; but scientific good sense concludes in many situations that it would be perfectly unreasonable to hold on to particular statements. (B obviously follows the Duhem thesis rather than the Quine thesis.)

. . . The thesis seems to me to be both true and important. . . .

■ | Notes

1. 1 degree = 60', and 1' = 60". So 52' is slightly less than a degree.
2. Einstein may have been influenced by Duhem, however, as is suggested by Howard in his interesting 1990 article. Howard shows that Einstein was on very friendly terms with Friedrich Adler, who prepared the first German translation of the *Aim and Structure of Physical Theory*, which appeared in 1908. From the autumn of 1909, Einstein and his wife rented an apartment in Zurich just immediately upstairs from the Adlers, and Einstein and Adler would meet frequently to discuss philosophy and physics. So probably Einstein had read *Aim and Structure* by the end of 1909 at the latest.
3. Or rather, misquotes. Duhem writes: 'raisons que la raison ne connaît pas' (1904–5, French edn, p. 330), whereas Pascal's original *pensée* was 'Le coeur a ses raisons que la raison ne connaît point.' Giving quotations which are slightly wrong is often a sign of great familiarity with a particular author.
4. Vuillemin (1979) and Ariew (1984) give valuable discussions of the differences between the Duhem thesis and the Quine thesis. I found these articles very helpful when forming my own views on the subject.
5. Quine's views have altered over the years, but here we will discuss only the position found in his 1951 article.
6. It is possible, however, to use arguments not involving the Quine thesis against the analytic/synthetic distinction. I give two such arguments against the distinction, the argument from justification and the argument from truth, in my 1985 article.

■ | References

- Ariew, R. 1984. "The Duhem Thesis," *British Journal for the Philosophy of Science* 35: 313–25.
- Crowe, M. J. 1990. "Duhem and the History and Philosophy of Mathematics," *Synthese* 83: 431–47.
- Duhem, P. 1904–5. *The Aim and Structure of Physical Theory*. English translation by Philip P. Wiener of the 2nd French edn. of 1914, Atheneum, 1962. French edn., Vrin, 1989.
- Duhem, P. 1915. *La Science allemande*. A. Hermann et Fils.
- Gillies, D. A. 1985. "The Analytic/Synthetic Problem," *Ratio* 27: 149–59.
- Howard, D. 1990. "Einstein and Duhem," *Synthese* 83:363–84.
- Martin, R. N. D. 1991. *Pierre Duhem. Philosophy and History in the Work of a Believing Physicist*. Open Court.
- Poincaré, H. 1902. *Science and Hypothesis*. English translation, Dover, 1952. French edn., Flammarion, 1968.

Poincaré, H. 1904. "L'État actuel et l'avenir de la physique mathématique." Lecture delivered 24 Sept. 1904 to the International Congress of Arts and Science, St. Louis, Missouri, and published in *Bulletin des sciences mathématiques* 28: 302-24. Reprinted in Poincaré, 1905, 91-111.

Poincaré, H. 1905. *The Value of Science*. English translation, Dover, 1958.

Quine, W. V. O. 1951. "Two Dogmas of Empiricism." Reprinted in *From a Logical Point of View*, 2nd rev. edn., Harper Torchbooks, 1961, 20-46.

Vuillemin, J. 1968. Préface to H. Poincaré, *La Science et l'hypothèse*, Flammarion, 1968, 7-19.

Vuillemin, J. 1979. "On Duhem's and Quine's Theses," *Grazer Philosophische Studien* 9: 69-96. Quotations are from reprint in *The Philosophy of W. V. Quine*, ed. L. E. Hahn and P. A. Schilpp, Library of Living Philosophers (La Salle, Ill.: Open Court, 1986), 595-618.

Demystifying Underdetermination

Pure logic is not the only rule for our judgments; certain opinions which do not fall under the hammer of the principle of contradiction are in any case perfectly unreasonable.

— Pierre Duhem¹

■ | Introduction

This essay begins with some good sense from Pierre Duhem. The piece can be described as a defense of this particular Duhemian thesis against a rather more familiar doctrine to which Duhem's name has often been attached. To put it in a nutshell, I shall be seeking to show that the doctrine of underdetermination, and the assaults on methodology that have been mounted in its name, founder precisely because they suppose that the logically possible and the reasonable are coextensive. Specifically, they rest on the assumption that, unless we can show that a scientific hypothesis cannot possibly be reconciled with the evidence, then we have no epistemic grounds for faulting those who espouse that hypothesis. Stated so baldly, this appears to be an absurd claim. That in itself is hardly decisive, since many philosophical (and scientific) theses smack initially of the absurd. But, as I shall show below in some detail, the surface implausibility of this doctrine gives way on further analysis to the conviction that it is even more untoward and ill argued than it initially appears. And what compounds the crime is that precisely this thesis is presupposed by many of the fashionable epistemologies of science of the last quarter cen-

FROM C. Wade Savage, ed., *Scientific Theories*, vol. 14, *Minnesota Studies in the Philosophy of Science* (Minneapolis: University of Minnesota Press, 1990), 267–97.

tury. Before this complex indictment can be made plausible, however, there is a larger story that has to be told.

There is abroad in the land a growing suspicion about the viability of scientific methodology. Polanyi, Wittgenstein, Feyerabend and a host of others have doubted, occasionally even denied, that science is or should be a rule-governed activity. Others, while granting that there are rules of the 'game' of science, doubt that those rules do much to delimit choice (e.g., Quine, Kuhn). Much of the present uneasiness about the viability of methodology and normative epistemology can be traced to a series of arguments arising out of what is usually called "the underdetermination of theories." Indeed, on the strength of one or another variant of the thesis of underdetermination, a motley coalition of philosophers and sociologists has drawn some dire morals for the epistemological enterprise.

Consider a few of the better-known examples: Quine has claimed that theories are so radically underdetermined by the data that a scientist can, if he wishes, hold on to any theory he likes, "come what may." Lakatos and Feyerabend have taken the underdetermination of theories to justify the claim that the only difference between empirically successful and empirically unsuccessful theories lay in the talents and resources of their respective advocates (i.e., with sufficient ingenuity, more or less any theory can be made to look methodologically respectable).² Boyd and Newton-Smith suggest that underdetermination poses several *prima facie* challenges to scientific realism.³ Hesse and Bloor have claimed that underdetermination shows the necessity for bringing noncognitive, social factors into play in explaining the theory choices of scientists (on the grounds that methodological and evidential considerations alone are demonstrably insufficient to account for such choices).⁴ H. M. Collins, and several of his fellow sociologists of knowledge, have asserted that underdetermination lends credence to the view that the world does little if anything to shape or constrain our beliefs about it.⁵ Further afield, literary theorists like Derrida have utilized underdetermination as one part of the rationale for "deconstructionism" (in brief, the thesis that, since every text lends itself to a variety of interpretations and thus since texts underdetermine choice among those interpretations, texts have no determinate meaning).⁶ This litany of invocations of underdeterminationist assumptions could be expanded almost indefinitely; but that is hardly called for, since it has become a familiar feature of contemporary intellectual discourse to endow underdetermination with a deep significance for our understanding of the limitations of methodology, and thus with broad ramifications for all our claims to knowledge—insofar as the latter are alleged to be grounded in trustworthy procedures of inquiry. In fact, underdetermination forms the central weapon in the relativistic assault on epistemology.

As my title suggests, I think that this issue has been overplayed. Sloppy formulations of the thesis of underdetermination have encouraged authors to use it—sometimes inadvertently, sometimes willfully—to support what-

ever relativist conclusions they fancy. Moreover, a failure to distinguish several distinct species of underdetermination—some probably viable, others decidedly not—has encouraged writers to lump together situations that ought to be sharply distinguished. Above all, inferences have been drawn from the fact of underdetermination that by no means follow from it. Because all that is so, we need to get as clear as we can about this slippery concept before we can decide whether underdetermination warrants the critiques of methodology that have been mounted in its name. That is the object of the next section of this paper. With those clarifications in hand, I will then turn in succeeding parts to assess some recent garden-variety claims about the methodological and epistemic significance of underdetermination.

Although this paper is one of a series whose larger target is epistemic relativism in general⁷, my limited aim here is not to refute relativism in all its forms. It is rather to show that one important line of argument, beloved of relativists, the argument from underdetermination, will not sustain the global conclusions that they claim to derive from it.

■ | Vintage Versions of Underdetermination

HUMEAN UNDERDETERMINATION

Although claims about underdetermination have been made for almost every aspect of science, those that interest philosophers most have to do specifically with claims about the underdetermination of *theories*. I shall use the term “theory” merely to refer to any set of *universal statements* that purport to describe the natural world.⁸ Moreover, so as not to make the underdeterminationists’ case any harder to make out than it already is, I shall—for purposes of this essay—suppose, with them, that single theories by themselves make no directly testable assertions. More or less everyone, relativist or nonrelativist, agrees that “theories are underdetermined” in some sense or other; but the seeming agreement about that formula disguises a dangerously wide variety of different meanings.

Our first step in trying to make some sense of the huge literature on underdetermination comes with the realization that there are two quite distinct families of theses, both of which are passed off as “the” thesis of underdetermination. Within each of these “families,” there are still further differentiating features. The generic and specific differences between these versions, as we shall see shortly, are not minor or esoteric. They assert different things; they presuppose different things; the arguments that lead to and from them are quite different. Nonetheless each has been characterized, and often, as “the doctrine of underdetermination.”

The first of the two generic types of underdetermination is what I shall call, for obvious reasons, deductive or *Humean underdetermination*

(HUD). It amounts to one variant or other of the following claim:

HUD For any finite body of evidence, there are indefinitely many mutually contrary theories, each of which logically entails that evidence.

The arguments for HUD are sufficiently familiar and sufficiently trivial that they need no rehearsal here.* HUD shows that the fallacy of affirming the consequent is indeed a deductive fallacy (like so many other interesting patterns of inference in science); that the method of hypothesis is not logically probative; that successfully "saving the phenomena" is not a robust warrant for detachment or belief. I have no quarrels with either HUD or with the familiar arguments that can be marshaled for it. But when duly considered, HUD turns out to be an extraordinarily weak thesis about scientific inference, one that will scarcely sustain any of the grandiose claims that have been made on behalf of underdetermination.

Specifically, HUD is weak in two key respects: First, it addresses itself only to the role of *deductive logic* in scientific inference; it is wholly silent about whether the rules of a broader ampliative logic underdetermine theory choice. Secondly, HUD provides no motivation for the claim that *all* theories are reconcilable with any given body of evidence: it asserts rather that indefinitely many theories are so. Put differently, even if our doxastic policies were so lax that they permitted us to accept as rational any belief that logically entailed the evidence, HUD would not sanction the claim (which we might call the "*thesis of cognitive egalitarianism*") that all rival theories are thereby equally belief-worthy or equally rational to accept.

Despite these crucial and sometimes overlooked limitations of its scope, HUD still has some important lessons for us. For instance, HUD makes clear that theories cannot be "deduced from the phenomena" (in the literal, non-Newtonian sense of that phrase). It thus establishes that the resources of deductive logic are insufficient, no matter how extensive the evidence, to enable one to determine for certain that any theory is true. But for anyone comfortable with the nowadays familiar mixture of (a) fallibilism about knowledge and (b) the belief that ampliative inference depends on modes of argument that go beyond deductive logic, none of that is either very surprising or very troubling.

As already noted, HUD manifestly does not establish that all theories are equally good or equally well supported, or that falsifications are inconclusive or that any theory can be held on to, come what may. Nor, finally, does it suggest, let alone entail, that the methodological enterprise

* For a rehearsal of some trivial arguments for HUD, see the section "Humean Underdetermination" in the commentary on chapter 3.

is hopelessly flawed because methodological rules radically underdetermine theory selection. Indeed, consistently with HUD, one could hold (although I shall not) that the ampliative rules of scientific method fully determine theory choice. HUD says nothing whatever about whether ampliative rules of theory appraisal do or do not determine theory choice uniquely. What HUD teaches, and all that it licenses, is that if one is prepared to accept only those theories that can be proven to be true, then one is going to have a drastically limited doxastic repertoire.

Mindful of the some of the dire consequences (enumerated above) that several authors have drawn from the thesis of underdetermination, one is inclined to invoke minimal charity by saying that Humean underdetermination must not be quite what they have in mind. And I think we have independent evidence that they do not. I have dwelt on this weak form of underdetermination to start with because, as I shall try to show below, it is the only *general* form of underdetermination that has been incontrovertibly established. Typically, however, advocates of underdetermination have a much stronger thesis in mind. Interestingly, when attacked, they often fall back on the truism of HUD; a safe strategy since HUD is unexceptionable. They generally fail to point out that HUD will support none of the conclusions that they wish to draw from underdetermination. By failing to distinguish between HUD and stronger (and more controversial) forms of underdetermination, advocates of undifferentiated underdetermination thus piggyback their stronger claims on this weaker one. But more of that below.

THE QUINEAN REFORMULATIONS OF UNDERDETERMINATION⁹

Like most philosophers, Quine of course accepts the soundness of HUD. But where HUD was silent on the key question of ampliative underdetermination, Quine (along with several other philosophers) was quick to take up the slack. In particular, Quine has propounded two distinct doctrines, both of which have direct bearing on the issues before us. The first, and weaker, of these doctrines I shall call the *nonuniqueness thesis*. It holds that: *for any theory, T, and any given body of evidence supporting T, there is at least one rival (i.e. contrary) to T that is as well supported as T.*¹⁰ In his more ambitious (and more influential) moments, Quine is committed to a much stronger position, which I call the *egalitarian thesis*. It insists that: *every theory is as well supported by the evidence as any of its rivals.*¹¹ Quine nowhere explicitly expresses the egalitarian thesis in precisely this form. But it will be the burden of the following analysis to show that Quine's numerous pronouncements on the retainability of theories, in the face of virtually any evidence, presuppose the egalitarian thesis, and make no sense without it. What follows is not meant to be an exegesis of Quine's intentions; it is meant, rather, as an exploration of whether Quine's posi-

tion on this issue will sustain the broad implications that many writers (sometimes including Quine himself) draw from it.

What distinguishes both the nonuniqueness thesis and the egalitarian thesis from HUD is that they concern ampliative rather than deductive underdetermination; that is, they centrally involve the notion of "empirical support," which is after all the central focus of ampliative inference. In this section and the first part of the next, I shall focus on Quine's discussion of these two forms of ampliative underdetermination (especially the egalitarian thesis), and explore some of their implications. The egalitarian thesis is sufficiently extreme—not to say epistemically pernicious—that I want to take some time showing that some versions of Quine's holism are indeed committed to it. I shall thus examine its status in considerable detail before turning in later sections to look at some other prominent accounts of ampliative underdetermination.

Everyone knows that Quine, in his "Two Dogmas of Empiricism," maintained that:

- (0) one may hold onto any theory whatever in the face of any evidence whatever.¹²

Crucial here is the sense of "may" involved in this extraordinary claim. If taken as asserting that human beings are psychologically capable of retaining beliefs in the face of overwhelming evidence against them, then it is a wholly uninteresting truism, borne out by every chapter in the saga of human folly. But if Quine's claim is to have any bite, or any philosophical interest, it must be glossed along roughly the following lines:

- (1) It is rational to hold onto any theory whatever in the face of any evidence whatever.

I suggest this gloss because I suppose that Quine means to be telling us something about scientific rationality; and it is clear that (0), construed descriptively, has no implications for normative epistemology. Combined with Quine's counterpart claim that one is also free to jettison any theory one is minded to, (1) appears to assert the *equivarationality* of all rival theoretical systems. Now, what grounds does Quine have for asserting (1)? One might expect that he could establish the plausibility of (1) only in virtue of examining the relevant rules of rational theory choice and showing, if it could be shown, that those rules were always so ambiguous that, confronted with any pair of theories and any body of evidence, they could never yield a decision procedure for making a choice. Such a proof, if forthcoming, would immediately undercut virtually every theory of empirical or scientific rationality. But Quine *nowhere*, neither in "Two Dog-

mas . . . " nor elsewhere, engages in a general examination of ampliative rules of theory choice.

His specific aim in propounding (0) or (1) is often said to be to exhibit the ambiguity of falsification or of *modus tollens*. The usual reading of Quine here is that he has shown the impotence of negative instances to disprove a theory, just as Hume had earlier showed the impotence of positive instances to prove a theory. Indeed, it is this gloss that establishes the parallel between Quine's form of the thesis of underdetermination and HUD. Between them, they seem to lay to rest any prospect for a purely deductive logic of scientific inference.

But what is the status of (1)? I have already said that Quine nowhere engages in an exhaustive examination of various rules of rational theory choice with a view to showing them impotent to make a choice between all pairs of theories. Instead, he is content to examine a *single* rule of theory choice, what we might call the Popperian gambit. That rule says, in effect, "reject theories that have (known) falsifying instances." Quine's strategy is to show that this particular rule radically underdetermines theory choice. I intend to spend the bulk of this section examining Quine's case for the claim that this particular rule underdetermines theory choice. But the reader should bear in mind that even if Quine were successful in his dissection of this particular rule (which he is not), that would still leave unsettled the question whether other ampliative rules of detachment suffer a similar fate.*

How does he go about exhibiting the underdeterminative character of falsification? Well, Quine's explicit arguments for (1) in "Two Dogmas . . ." are decidedly curious. Confronted, for instance, with an apparent refutation of a claim that "there are brick houses on Elm Street," we can—he says—change the meaning of the terms so that (say) "Elm Street" now refers to Oak Street, which adventitiously happens to have brick houses on it, thereby avoiding the force of the apparent refutation. Now this is surely a Pickwickian sense of "holding onto a theory come what may," since what we are holding onto here is not what the theory asserted, but the (redefined) string of words constituting the theory.¹³ Alternatively, says Quine, we can always change the laws of logic if need be. We might, one supposes, abandon *modus tollens*, thus enabling us to maintain a theory in the face of evidence that, under a former logical regime, was falsifying of it; or we could jettison *modus ponens* and thereby preclude the possibility that the theory we are concerned to save is "implicated" in any schema of inference leading to the awkward prediction. If one is loath to abandon such useful logical devices (and Quine is), other resources are

* In inductive logic, rules of detachment are often called *acceptance rules*. They specify when it is permissible to accept a hypothesis as true, thus "detaching" the hypothesis h from the assertion $P(h/e) = r$, that h has probability r on evidence e .

open to us. We could, says Quine, dismiss the threatening evidence "by pleading hallucination."¹⁴

But are there no constraints on when it is reasonable to abandon selected rules of logic or when to label evidence specious (because the result of hallucination) or when to redefine the terms of our theories? Of course, it is (for all I know) humanly possible to resort to any of these stratagems, as a descriptivist reading of (0) might suggest. But nothing Quine has said thus far gives us any grounds to believe, as (1) asserts, that it will ever, let alone *always*, be rational to do so. Yet his version of the thesis of underdetermination, if he means it to have any implications for normative epistemology, requires him to hold that it is rational to use some such devices.¹⁵ Hence he would appear to be committed to the view that epistemic rationality gives us no grounds for avoiding such maneuvers. (On Quine's view, the only considerations that we could possibly invoke to block such stratagems have to do with pragmatic, not epistemic, rationality.¹⁶) Thus far, the argument for ampliative underdetermination seems made of pretty trifling stuff.

But there is a fourth, and decidedly nontrivial, stratagem that Quine envisages for showing how our Popperian principle underdetermines theory choice. This is the one that has received virtually all the exegetical attention: quite rightly too, since Quine's arguments on the other three are transparently question begging because they fail to establish the rationality of holding onto any theory in the face of any evidence. Specifically, Quine proposes that a threatened statement or theory can always be immunized from the threat of the recalcitrant evidence by making suitable adjustments in our auxiliary theories. It is here that the familiar "Duhem-Quine thesis" comes to the fore. What confronts experience in any test, according to both Quine and Duhem, is an entire theoretical structure (later dubbed by Quine "a web of belief") consisting inter alia [among other things] of a variety of theories. Predictions, they claim, can never be derived from single theories but only from collectives consisting of multiple theories, statements of initial and boundary conditions, assumptions about instrumentation, and the like. Since (they claim) it is whole systems and whole systems alone that make predictions, when those predictions go awry it is theory complexes, not individual theories, that are indicted via *modus tollens*. But, so the argument continues, we cannot via *modus tollens* deduce the falsity of any component of a complex from the falsity of the complex as a whole. Quine put it this way:

But the failure [of a prediction] falsifies only a block of theory as a whole, a conjunction of many statements. The failure shows that one or more of those statements is false, but it does not show which.¹⁷

Systems, complexes or "webs" apparently turn out to be unambiguously falsifiable on Quine's view; but the choice between individual theories

or statements making up these systems is, in his view, radically underdetermined.

Obviously, this approach is rather more interesting than Quine's other techniques for saving threatened theories, for here we need not abandon logic, redefine the terms in our theories in patently ad hoc fashion, nor plead hallucinations. The thesis of underdetermination in this particular guise, which I shall call Quinean underdetermination (QUD), can be formulated as follows:

QUD Any theory can be reconciled with any recalcitrant evidence by making suitable adjustments in our other assumptions about nature.

Before we comment on the credentials of QUD, we need to further disambiguate it. We especially need to focus on the troublesome phrase "can be reconciled with." On a weak interpretation, this would be glossed as "can be made logically compatible with the formerly recalcitrant evidence." I shall call this the *"compatibilist version of QUD."* On a stronger interpretation, it might be glossed as "can be made to function significantly in a complex that entails" the previously threatening evidence. Let us call this the *"entailment version of QUD."* To repeat, the compatibilist version says that any theory can be made *logically compatible* with any formerly threatening evidential report; the entailment interpretation insists further that any theory can be made to function essentially in a *logical derivation* of the erstwhile refuting instance.

The compatibilist version of QUD can be trivially proven. All we need do, given any web of belief and a suspect theory that is part of it, is to remove (*without replacement*) any of those ancillary statements within the web needed to derive the recalcitrant prediction from the theory. Of course, we may well lose enormous explanatory power thereby, and the web may lose much of its pragmatic utility thereby, but there is nothing in deductive logic that would preclude any of that.

The entailment version of QUD, by contrast, insists that there is always a set of auxiliary assumptions that can replace others formerly present, and that will allow the *derivation*, not of the wrongly predicted result, but of precisely what we have observed. As Grünbaum, Quinn, Laudan and others have shown,¹⁸ neither Quine nor anyone else has ever produced a general existence proof concerning the availability either in principle or in practice of suitable (i.e., nontrivial) theory-saving auxiliaries. Hence the entailment version of QUD is without apparent warrant. For a time (circa 1962), Quine himself conceded as much.¹⁹ That is by now a familiar result. But what I think needs much greater emphasis than it has received is the fact that, *even if nontrivial auxiliaries existed that would satisfy the demands of the entailment version of QUD, no one has ever shown that it would be rational to prefer a web that included them and the threatened*

theory to a rival web that dispensed with the theory in question. Indeed, as I shall show in detail, what undermines *both* versions of QUD is that neither logical compatibility with the evidence nor logical derivability of the evidence is sufficient to establish that a theory exhibiting such empirical compatibility and derivability is rationally acceptable.

It will prove helpful to distinguish four different positive relations in which a theory (or the system in which a theory is embedded) can stand to the evidence. Specifically, a theory (or larger system of which it is a part) may:

- be logically compatible with the evidence;
- logically entail the evidence;
- explain the evidence;
- be empirically supported by the evidence.

Arguably, none of these relations reduces to any of the others; despite that, Quine's analysis runs all four together. But what is especially important for our purposes is the realization that *satisfaction of either the compatibility relation or the entailment relation fails to establish either an explanatory relation or a relation of empirical support.* For instance, theories may entail statements that they nonetheless do not explain; self-entailment being the most obvious example. Equally, theories may entail evidence statements, yet not be empirically supported by them (e.g., if the theory was generated by the algorithmic manipulation of the "evidence" in question).

So, when QUD tells us that any theory can be "reconciled" with any bit of recalcitrant evidence, we are going to have to attend with some care to what that reconciliation consists in. Is Quine claiming, for instance, that any theory can—by suitable modifications elsewhere—continue to function as part of an *explanation* of a formerly recalcitrant fact? Or is he claiming, even more ambitiously, that any formerly recalcitrant instance for a theory can be transformed into a *confirming instance* for it?

As we have seen, the only form of QUD that has been firmly established is compatibilist Quinean underdetermination (an interpretation that says a theory can always be rendered logically compatible with any evidence, provided we are prepared to give up enough of our other beliefs); so I shall begin my discussion there. Saving a prized, but threatened, theory by abandoning the auxiliary assumptions once needed to link it with recalcitrant evidence clearly comes at a price. Assuming that we give up those beliefs without replacement (and recall that this is the only case that has been made plausible), we not only abandon an ability to say anything whatever about the phenomena that produced the recalcitrant experience; we also now give up the ability to explain all the other things which those now-rejected auxiliaries enabled us to give an account of—

with no guarantee whatever that we can find alternatives to them that will match their explanatory scope.

But further and deeper troubles lurk for Quine just around the corner. For it is not just explanatory scope that is lost; it is also *evidential support*. Many of those phenomena that our web of belief could once give an account of (and which presumably provided part of the good reasons for accepting the web with its constituent theories) are now beyond the resources of the web to explain and predict. That is another way of saying that the revised web, stripped of those statements formerly linking the theory in question with the mistaken prediction, now has substantially less empirical support than it once did; assuming, of course, that the jettisoned statements formerly functioned to do more work for us than just producing the discredited prediction.²⁰ Which clearly takes things from bad to worse. For now Quine's claim about the salvageability of a threatened theory turns out to make sense just in case the only criterion of theory appraisal is logical compatibility with observation. If we are concerned with issues like explanatory scope or empirical support, Quine's QUD in its compatibilist version cuts no ice whatsoever.

Clearly, what is wrong with QUD, and why it fails to capture the spirit of (1), is that it has dropped out any reference to the *rationality* of theory choices, and specifically theory rejections. It doubtless is possible for us to jettison a whole load of auxiliaries in order to save a threatened theory (where "save" now means specifically "to make it logically compatible with the evidence"), but Quine nowhere establishes the reasonableness or the rationality of doing so. And if it is plausible, as I believe it is, to hold that scientists are aiming (among other things) at producing theories with broad explanatory scope and impressive empirical credentials, then it has to be said that Quine has given us no arguments to suppose that any theory we like can be doctored up so as to win high marks on those scores.

This point underscores the fact that too many of the discussions of underdetermination in the last quarter century have proceeded in an evaluative vacuum. They imagine that if a course of action is logically possible, then one need not attend to the question of its rationality. But if QUD is to carry any epistemic force, it needs to be formulated in terms of the rationality of preserving threatened theories. One might therefore suggest the following substitute for QUD (which was itself a clarification of (1)):

- (2) any theory can be rationally retained in the face of any recalcitrant evidence.

Absent strong arguments for (2) or its functional equivalents, Quinean holism, the Duhem-Quine thesis and the (non-Humean) forms of underdetermination appear to pose no threat in principle for an account of

scientific methodology or rationality. The key question is whether Quine, or any of the other influential advocates of the methodological significance of underdetermination, have such arguments to make.

Before we attempt to answer that question, a bit more clarification is called for, since the notion of retainment, let alone rational retainment, is still less than transparent. I propose that we understand that phrase to mean something along these lines: to say that a theory can be rationally retained is to say that reasons can be given for holding that theory, or the system of which it is a part, as true (or empirically adequate) that are (preferably stronger than but) as least as strong as the reasons that can be given for holding as true (or empirically adequate) any of its *known* rivals. Some would wish to give this phrase a more demanding gloss; they would want to insist that a theory can be rationally held only if we can show that the reasons in its behalf are stronger than those for all its *possible* rivals, both extant and those yet-to-be-conceived. That stronger gloss, which I shall resist subscribing to, would have the effect of making it even harder for Quine to establish (2) than my weaker interpretation does. Because I believe that theory choice is generally a matter of comparative choice among extant alternatives, I see no reason why we should saddle Quine and his followers with having to defend (2) on its logically stronger construal. More to the point, if I can show that the arguments on behalf of the weaker construal fail, that indeed the weaker construal is false, it follows that its stronger counterpart fails as well, since the stronger entails the weaker. I therefore propose emending (2) as follows:

- (2*) any theory can be shown to be as well supported by any evidence as any of its known rivals.

Quine never formulates this thesis as such, but I have tried to show that defending a thesis of this sort is incumbent on anyone who holds, as Quine does, that any theory can be held true, come what may. Duly considered, (2*) is quite a remarkable thesis, entailing as it does that all the known contraries to every known theory are equally well supported. Moreover, (2*) is our old friend, the egalitarian thesis. If correct, (2*) entails (for instance) that the flat-earth hypothesis is as sound as the oblate-spheroid hypothesis²¹; that it is as reasonable to believe in fairies at the bottom of my garden as not. But, for all its counter-intuitiveness, this is precisely the doctrine to which authors like Quine, Kuhn, and Hesse are committed.²² (In saying that Quine is committed to this position, I do not mean that he would avow it if put to him directly; I doubt that very much. My claim rather is (a), that Quine's argument in "Two Dogmas . . ." commits him to such a thesis, and (b), that those strong relativists who look to Quine as having espoused and established the egalitarian thesis are exactly half right. I prefer to leave it to Quine exegetes to decide whether the positions

of the *later* Quine allow him to be exonerated of the charge that his more recent writing run afoul of the same problem.)

One looks in vain in "Two Dogmas . . ." for even the whiff of an argument that would make the egalitarian thesis plausible. As we have seen, Quine's only marginally relevant points there are his suppositions (1) that any theory can be made logically compatible with any evidence (statement) and (2) that any theory can function in a network of statements that will entail any particular evidence statement.²³ But what serious epistemologist has ever held either (a) that bare logical compatibility with the evidence constituted adequate reason to accept a scientific theory,²⁴ or (b) that logical entailment of the evidence by a theory constituted adequate grounds for accepting a theory? One might guess otherwise. One might imagine that some brash hypothetico-deductivist would say that any theory that logically entailed the known evidence was acceptable. If one conjoins this doctrine with Quine's claim (albeit one that Quine has never made out) that every theory can be made to logically entail any evidence, then one has the makings of the egalitarian thesis. But such musings cut little ice, since no serious twentieth-century methodologist has ever espoused, without crucial qualifications, logical compatibility with the evidence or logical derivability of the evidence as a sufficient condition for detachment of a theory.²⁵

Consider some familiar theories of evidence to see that this is so. Within Popper's epistemology, two theories, T_1 and T_2 , that thus far have the same positive instances, e , may nonetheless be differentially supported by e . For instance, if T_1 predicted e before e was determined to be true, whereas T_2 is produced after e is known, then e (according to Popper) constitutes a good test of T_1 but no test of T_2 . Bayesians too insist that rival (but nonequivalent) theories sharing the same known positive instances are not necessarily equally well confirmed by those instances. Indeed, if two theories begin with different prior probabilities, then their posterior probabilities must be different, *given the same positive instances*.²⁶ But that is just to say that even if two theories enjoy precisely the same set of known confirming instances, *it does not follow that they should be regarded as equally well confirmed by those instances*. All of which is to say that showing that rival theories enjoy the same "empirical support"—in any sense of that term countenanced by (2*)—requires more than that those rivals are compatible with, or capable of entailing, the same "supporting" evidence. (2*) turns out centrally to be a claim in the theory of evidence and, since Quine does not address the evidence relation in "Two Dogmas . . .," one will not find further clarification of this issue there.²⁷

Of course, "Two Dogmas . . ." was not Quine's last effort to grapple with these issues. Some of these themes recur prominently in *Word and Object*, and it is worth examining some of Quine's arguments about underdetermination to be found there. In that work, Quine explicitly if briefly addresses the question, already implicit in "Two Dogmas . . .,"

whether ampliative rules of theory choice underdetermine theory choice.³⁸ Quine begins his discussion there by making the relatively mild claim: that scientific methodology, along with any imaginable body of evidence, might possibly underdetermine theory choice. As he wrote:

*conceivably the truths about molecules are only partially determined by any ideal organon of scientific method plus all the truths that can be said in common sense terms about ordinary things.*³⁹

Literally, the remark in this passage is unexceptionable. Since we do not yet know what the final "organon of scientific method" will look like, it surely is "conceivable" that the truth status of claims about molecular structure might be underdetermined by such an organon. Three sentences later, however, this claim about the conceivability of ampliative underdetermination becomes a more ambitious assertion about the *likelihood* of such underdetermination:

*The incompleteness of determination of molecular behavior by the behavior of ordinary things . . . remains true even if we include all past, present and future irritations of all the far-flung surfaces of mankind, and probably even if we throw in [i.e., take for granted] an in fact achieved organon of scientific method besides.*⁴⁰

As it stands, and as it remains in Quine's text, this is no argument at all, but a bare assertion. But it is one to which Quine returns still later:

we have no reason to suppose that man's surface irritations even unto eternity admit of any systematization that is scientifically better or simpler than all possible others. It seems likelier, if only on account of symmetries or dualities, that countless alternative theories would be tied for first place."

Quite how Quine thinks he can justify this claim of "likelihood" for ampliative underdetermination is left opaque. Neither here nor elsewhere does he show that any specific ampliative rules of scientific method⁴¹ actually underdetermine theory choice—let alone that the rules of a "final methodology" will similarly do so. Instead, on the strength of the notorious ambiguities of simplicity (and by some hand-waving assertions that other principles of method may "plausibly be subsumed under the demand for simplicity")⁴²—a claim that is anything but plausible), Quine asserts "in principle," that there is "probably" no theory that can uniquely satisfy the "canons of any ideal organon of scientific method."⁴³ In sum, Quine fails to show that theory choice is ampliatively underdetermined even by *existing* codifications of scientific methodology (all of which go considerably beyond the principle of simplicity), let alone by all possible such codifications.⁴⁴

More important for our purposes, even if Quine were right that no ideal organon of methodology could ever pick out any theory as uniquely satisfying its demands, we should note—in the version of underdetermination contained in the last passage from Quine—how drastically he has apparently weakened his claims from those of “Two Dogmas. . . .” That essay, you recall, had espoused the egalitarian thesis that *any* theory can be reconciled with any evidence. We noted how much stronger that thesis was than the nonuniqueness thesis to the effect that there will always be some rival theories reconcilable with any finite body of evidence. But in *Word and Object*, as the passages I have cited vividly illustrate, Quine is no longer arguing that *any* theory can be reconciled with *any* evidence;³⁶ he is maintaining rather that, no matter what our evidence and no matter what our rules of appraisal, there will always remain the possibility (or the likelihood) that the choice will not be uniquely determined. But that is simply to say that there will (probably) always be at least one contrary to any given theory that fits the data equally well—a far cry from the claim, associated with QUD and (2*), that *all* the contraries to a given theory will fit the data equally well. In a sense, therefore, Quine appears in *Word and Object* to have abandoned the egalitarian thesis for the nonuniqueness thesis, since the latter asserts not the epistemic equality of all theories but only the epistemic equality of certain theories.³⁷ That surmise aside, it is fair to say that *Word and Object* does nothing to further the case for Quine’s egalitarian view that “any theory can be held true come what may.”

Some terminological codification might be useful before we proceed, since we have reached a natural breaking point in the argument. As we have seen, one can distinguish between (a) *descriptive* (0) and (b) *normative* (1, 2, 2*) forms of underdetermination, depending upon whether one is making a claim about what people are capable of doing or what the rules of scientific rationality allow.³⁸ One can also distinguish between (c) *deductive* and (d) *ampliative* underdetermination, depending upon whether it is the rules of deductive logic (HUD) or of a broadly inductive logic or theory of rationality that are alleged to underdetermine choice (QUD). Further, we can distinguish between the claims that theories can be reconciled with recalcitrant evidence via establishing (e) *compatibility* between the two or (f) a one-way *entailment* between the theory and the recalcitrant evidence or (g) *equivalence of support* between rival theories. Finally, one can distinguish between (h) the doctrine that choice is underdetermined between at least one of the contraries of a theory and that theory (*nonuniqueness*) and (i) the doctrine that theory choice is underdetermined between every contrary of a theory and that theory (“cognitive egalitarianism”).

Using this terminology, we can summarize such conclusions as we have reached to this point: In “Two Dogmas. . . .,” Quine propounded a thesis of normative, ampliative, egalitarian underdetermination. Whether

we construe that thesis in its compatibilist or entailment versions, it is clear that Quine has said nothing that makes plausible the idea that every *prima facie* refuted theory can be embedded in a rationally acceptable (i.e., empirically well-supported) network of beliefs. Moreover, "Two Dogmas . . ." developed an argument for underdetermination for only one rationality principle among many, what I have been calling the Popperian gambit. This left completely untouched the question whether other rules of theory choice suffered from the same defects that Quine thought Popper's did. Perhaps with a view to remedying that deficiency, Quine argued—or, rather, alleged without argument—in *Word and Object* that any codification of scientific method would underdetermine theory choice. Unfortunately, *Word and Object* nowhere delivers on its claim about underdetermination.

But suppose, just for a moment, that Quine had been able to show what he claimed in *Word and Object*, to wit, the nonuniqueness thesis. At best, that result would establish that for any well-confirmed theory, there is in principle at least one other theory that will be equally well-confirmed by the same evidence. That is an interesting thesis to be sure, and possibly a true one, although Quine has given us no reason to think so. (Shortly, we shall examine arguments of other authors that seem to provide some ammunition for this doctrine.) But even if true, the nonuniqueness thesis will not sustain the critiques of methodology that have been mounted in the name of underdetermination. Those critiques are all based, implicitly or explicitly, on the strong, egalitarian reading of underdetermination. They amount to saying that the project of developing a methodology of science is a waste of time since, no matter what rules of evidence we eventually produce, those rules will do nothing to delimit choice between rival theories. The charge that methodology is toothless pivots essentially on the viability of QUD in its ampliative, egalitarian version. Nonuniqueness versions of the thesis of ampliative underdetermination at best establish that methodology will not allow us to pick out a theory as uniquely true, no matter how strong its evidential support. (*Word and Object's* weak ampliative thesis of underdetermination, even if sound, would provide no grounds for espousing the strong underdeterminationist thesis implied by the "any theory can be held come what may" dogma.³⁰)

Theory choice may or may not be ampliatively underdetermined in the sense of the nonuniqueness thesis; that is an open question. But however that issue is resolved, that form of underdetermination poses no challenge to the methodological enterprise. What would be threatening to, indeed debilitating for, the methodological enterprise is if QUD in its egalitarian version were once established. Even though Quine offers no persuasive arguments in favor of normative, egalitarian, ampliative underdetermination, there are several other philosophers who appear to have taken up the cudgels on behalf of precisely such a doctrine. It is time I turned to their arguments.

■ | Ampliative Underdetermination

With this preliminary spade work behind us, we are now in a position to see that the central question about underdetermination, at least so far as the philosophy of science is concerned, is the issue of ampliative underdetermination. Moreover, as we have seen, the threat to the epistemological project comes, not from the nonuniqueness version of underdetermination, but from the egalitarian version. (That version states that any theory can be embedded in a system that will be as strongly supported by the evidence as any rival is supported by the same evidence.) The question is whether anyone has stronger arguments than Quine's for the methodological underdetermination of theory choice. Two plausible contenders for that title are Nelson Goodman and Thomas Kuhn. I shall deal briefly with them in turn.

Goodman's *Fact, Fiction and Forecast* is notorious for posing a particularly vivid form of ampliative underdetermination, in the form of the *grue/green*, and related, paradoxes of induction.* Goodman is concerned there to deliver what Quine had elsewhere merely promised, namely, a proof that the inductive rules of scientific method underdetermine theory choice in the face of any conceivable evidence. The general structure of Goodman's argument is too familiar to need any summary here. But it is important to characterize carefully what Goodman's result shows. I shall do so utilizing terminology we have already been working with. Goodman shows that one specific rule of ampliative inference (actually a whole family of rules bearing structural similarities to the straight rule of induction) suffers from this defect: Given any pair (or *n*-tuple) of properties that have previously always occurred together in our experience, it is possible to construct an indefinitely large variety of contrary theories, all of which are compatible with the inductive rule: "If, for a large body of instances, the ratio of the successful instances of a hypothesis is very high compared to its failures, then assume that the hypothesis will continue to enjoy high

* Goodman defines the predicate *grue* in the following way: an object is *grue* if and only if it is observed before time *T* and is green, or else it is not observed before time *T* and is blue. Suppose we examine a large number of emeralds and find that all emeralds we have observed are green. We might reasonably infer that the hypothesis "All emeralds are green" is probably true (because it is confirmed by its many instances) and then use that hypothesis to predict that emeralds observed in the future (including those observed after *T*) will be green, too. But if our inspection has taken place before time *T* (say, the year 2050), then "All emeralds are *grue*" is also confirmed by the observed instances, and we could use the *grue* hypothesis to predict that any emerald observed after the year 2050 will be blue (and not green). Thus, we seem to have rival hypotheses, supporting incompatible predictions, both confirmed equally well by our evidence. For more on the *grue* problem and attempts to solve it, see "Goodman's Gruesome New Riddle of Induction" in the commentary on chapter 3.

success in the future." All these contraries will (along with suitable initial conditions) entail all the relevant past observations of the pairings of the properties in question. Thus, in one of Goodman's best-known examples, the straight rule will not yield an algorithm for choosing between "All emeralds are green" and "All emeralds are grue"; it awards them equally good marks.

There is some monumental question begging going on in Goodman's setting up of his examples. He supposes without argument that—since the contrary inductive extrapolations all have the same positive instances (to date)—the inductive logician must assume that the extrapolations from each of these hypotheses are all rendered equally likely by those instances. Yet we have already had occasion to remark that "possessing the same positive instances" and "being equally well confirmed" boil down to the same thing only in the logician's never-never land. (It was Whewell, Peirce and Popper who taught us all that theories sharing the same positive instances need not be regarded as equally well tested or equally belief-worthy.) But Goodman does have a point when he directs our attention to the fact that the straight rule of induction, as often stated, offers no grounds for distinguishing between the kind of empirical support enjoyed by the green hypothesis and that garnered by the grue hypothesis.

Goodman himself believes, of course, that this paradox of induction can be overcome by an account of the entrenchment of predicates. Regardless whether one accepts Goodman's approach to that issue, it should be said that strictly he does not hold that theory choice is underdetermined; on his view, such ampliative underdetermination obtains only if we limit our organon of scientific methodology to some version of the straight rule of induction.

But, for purposes of this paper, we can ignore the finer nuances of Goodman's argument since, even if a theory of entrenchment offered no way out of the paradox, and even if the slide from "possessing the same positive instances" to "being equally well confirmed" was greased by some plausible arguments, Goodman's arguments can provide scant comfort to the relativist's general repudiation of methodology. Recall that the relativist is committed, as we have seen, to arguing an egalitarian version of the thesis of ampliative underdetermination, i.e., he must show that all rival theories are equally well supported by any conceivable evidence. But there is nothing whatever in Goodman's analysis—even if we grant *all* its controversial premises—that could possibly sustain such an egalitarian conclusion. Goodman's argument, after all, does not even claim to show *a*propos of the straight rule that it will provide support for any and every hypothesis: his concern, rather, is to show that there will always be a family of contrary hypotheses between which it will provide no grounds for rational choice. The difference is crucial. If I propound the hypothesis that "All emeralds are red" and if my evidence base happens to be that all previously examined emeralds are green, then the straight rule is unambiguous in its

insistence that my hypothesis be rejected. The alleged inability of the straight rule to distinguish between ~~green~~ and ~~grae~~-style hypotheses provides no ammunition for the claim that such a rule can make no epistemic distinctions whatever between rival hypotheses. If we are confronted with a choice between (say) the hypotheses that all emeralds are red and that all are green, then the straight rule gives us entirely unambiguous advice concerning which is better supported by the relevant evidence. Goodmanian underdetermination is thus of the nonuniqueness sort. When one combines that with a recognition that Goodman has examined but one among a wide variety of ampliative principles that arguably play a role in scientific decision making, it becomes clear that no global conclusions whatever can be drawn from Goodman's analysis concerning the general inability of the rules of scientific methodology for strongly delimiting theory choice.

But we do not have to look very far afield to find someone who does propound a strong (viz., egalitarian) thesis of ampliative underdetermination, one which, if sound, would imply that the rules of methodology were never adequate to enable one to choose between any rival theories, regardless of the relevant evidence. I refer, of course, to Thomas Kuhn's assertion in *The Essential Tension* to the effect that the shared rules and standards of the scientific community *always* underdetermine theory choice.⁴ Kuhn there argues that science is guided by the use of several methods (or, as he prefers to call them, "standards"). These include the demand for empirical adequacy, consistency, simplicity, and the like. What Kuhn says about these standards is quite remarkable. He is not making the point that the later Quine and Goodman made about the methods of science; namely, that for any theory picked out by those methods, there will be indefinitely many contraries to it that are equally compatible with the standards. On the contrary, Kuhn is explicitly pushing the same line that the early Quine was implicitly committed to, viz., that the methods of science are inadequate ever to indicate that any theory is better than any rival, regardless of the available evidence. In the language of this essay, it is the egalitarian form of underdetermination that Kuhn is here proposing.

Kuhn, of course, does not use that language, but a brief rehearsal of Kuhn's general scheme will show that egalitarian underdetermination is one of its central underpinnings. Kuhn believes that there are divergent paradigms within the scientific community. Each paradigm comes to be associated with a particular set of practices and beliefs. Once a theory has been accepted within an ongoing scientific practice, Kuhn tells us, there is nothing that the shared standards of science can do to dislodge it. If paradigms do change, and Kuhn certainly believes that they do, this must be the result of "individual" and "subjective" decisions by individual researchers, not because there is anything about the methods or standards scientists share that ever requires the abandonment of those paradigms

and their associated theories. In a different vein, Kuhn tells us that a paradigm always looks good by its own standards, and weak by the standards of its rivals and that there never comes a point at which adherence to an old paradigm or resistance to a new one ever becomes "unscientific."⁴¹ In effect, then, Kuhn is offering a paraphrase of the early Quine, but giving it a Wittgensteinian twist: "once a theory/paradigm has been established within a practice, it can be held on to, come what may." The shared standards of the scientific community are allegedly impotent ever to force the abandonment of a paradigm, and the specific standards associated with any paradigm will always give it the nod.

If this seems extreme, I should let Kuhn speak for himself. "Every individual choice between competing theories," he tells us, "depends on a mixture of objective and subjective factors, or of shared and individual criteria."⁴² It is, in Kuhn's view, no accident that individual or subjective criteria are used alongside the objective or shared criteria, for the latter "are not by themselves sufficient to determine the decisions of individual scientists."⁴³ Each individual scientist "must complete the objective criteria [with 'subjective considerations'] before any computations can be done."⁴⁴ Kuhn is saying here that the shared methods or standards of scientific research are always insufficient to justify the choice of one theory over another.⁴⁵ That could only be so if (2") or one of its functional equivalents were true of those shared methods.

What arguments does Kuhn muster for this egalitarian claim? Well, he asserts that all the standards that scientists use are ambiguous and that "individuals may legitimately differ about their application to concrete cases."⁴⁶ "Simplicity, scope, fruitfulness and even accuracy can be judged differently . . . by different people."⁴⁷ He is surely right about some of this. Notoriously, one man's simplicity is another's complexity; one may think a new approach fruitful, while a second may see it as sterile. But such fuzziness of conception is precisely why most methodologists have avoided falling back on these hazy notions for talking about the empirical warrant for theories. Consider a different set of standards, one arguably more familiar to philosophers of science:

- prefer theories that are internally consistent;
- prefer theories that correctly make some predictions that are surprising given our background assumptions;
- prefer theories that have been tested against a diverse range of kinds of phenomena to those that have been tested only against very similar sorts of phenomena.

Even standards such as these have some fuzziness around the edges, but can anyone believe that, confronted with *any* pair of theories and *any* body of evidence, these standards are so rough-hewn that they could be

used indifferently to justify choosing either element of the pair? Do we really believe that Aristotle's physics correctly made the sorts of surprising predictions that Newton's physics did? Is there any doubt that Cartesian optics, with its dual insistence on the instantaneous propagation of light and that light traveled faster in denser media than in rarer ones, violated the canon of internal consistency?

Like the early Quine, Kuhn's wholesale holism commits him to the view that, consistently with the shared canons of rational acceptance, any theory or paradigm can be preserved in the face of any evidence. As it turns out, however, Kuhn no more has plausible arguments for this position than Quine had. In each case, the idea that the choice between changing or retaining a theory/paradigm is ultimately and always a matter of personal preference turns out to be an unargued dogma. In each case, if one takes away that dogma, much of the surrounding edifice collapses.

Of course, none of what I have said should be taken to deny that all forms of underdetermination are bogus. They manifestly are not. Indeed, there are several types of situations in which theory choice is indeed underdetermined by the relevant evidence and rules. Consider a few:

a) We can show that for some rules, and for certain theory pairs, theory choice is underdetermined for certain sorts of evidence. Consider the well-known case of the choice between the astronomical systems of Ptolemy and Copernicus. If the only sort of evidence available to us involves reports of line-of-sight positions of planetary position, and if our methodological rule is something like "Save the phenomena," then it is easy to prove that any line-of-sight observation that supports Copernican astronomy also supports Ptolemy's.⁴⁸ (It is crucial to add, of course, that if we consider other forms of evidence besides line-of-sight planetary position, this choice is not strongly underdetermined.)

b) We can show that for some rules and for some local situations, theory choice is underdetermined, regardless of the sorts of evidence available. Suppose our only rule of appraisal says, "Accept that theory with the largest set of confirming instances," and that we are confronted with two rival theories that have the same known confirming instances. Under these special circumstances, the choice is indeterminate.⁴⁹

What is the significance of such limited forms of ampliative underdetermination as these? They represent interesting cases to be sure, but none of them—taken either singly or in combination—establishes the soundness of strong ampliative underdetermination as a general doctrine. Absent sound arguments for global egalitarian underdetermination (i.e., afflicting every theory on every body of evidence), the recent dismissals of scientific methodology turn out to be nothing more than hollow, anti-intellectual sloganeering.

I have thus far been concerned to show that the case for strong ampliative underdetermination has not been convincingly made out. But we can more directly challenge it by showing its falsity in specific concrete

cases. To show that it is ill conceived (as opposed to merely unproved), we need to exhibit a methodological rule, or set of rules, a body of evidence, and a local theory choice context in which the rules and the evidence would *unambiguously* determine the theory preference. At the formal level it is of course child's play to produce a trivial rule that will unambiguously choose between a pair of theories. (Consider the rule: "Always prefer the later theory.") But, unlike the underdeterminationists,⁵⁰ I would prefer real examples, so as not to take refuge behind contrived cases.

The history of science presents us with a plethora of such cases. But I shall refer to only one example in detail, since that is all that is required to make the case. It involves the testing of the Newtonian celestial mechanics by measurements of the "bulging" of the earth.⁵¹ The Newtonian theory predicted that the rotation of the earth on its axis would cause a radical protrusion along the equator and a constriction at the poles—such that the earth's actual shape would be that of an oblate spheroid, rather than (as natural philosophers from Aristotle through Descartes had maintained) that of a uniform sphere or a sphere elongated along the polar axis. By the early eighteenth century, there were well-established geodesic techniques for ascertaining the shape and size of the earth (to which all parties agreed). These techniques involved the collection of precise measurements of distance from selected portions of the earth's surface. (To put it oversimplily, these techniques generally involved comparing measurements of chordal segments of the earth's polar and equatorial circumferences.⁵²) Advocates of the two major cosmogonies of the day, the Cartesians and the Newtonians, looked to such measurements as providing decisive evidence for choosing between the systems of Descartes and Newton.⁵³ At great expense, the Paris Académie des Sciences organized a series of elaborate expeditions to Peru and Lapland to collect the appropriate data. The evidence was assembled by scientists generally sympathetic to the Cartesian/Cassini hypothesis. Nonetheless, it was *their* interpretation, as well as everyone else's, that the evidence indicated that the diameter of the earth at its equator was significantly larger than along its polar axis. This result, in turn, was regarded as decisive evidence showing the superiority of Newtonian over Cartesian celestial mechanics. The operative methodological rule in the situation seems to have been something like this:

when two rival theories, T_1 and T_2 , make conflicting predictions that can be tested in a manner that presupposes neither T_1 nor T_2 , then one should accept whichever theory makes the correct prediction and reject its rival.

(I shall call this rule R_1 .) We need not concern ourselves here with whether R_1 is methodologically sound. The only issue is whether it underdetermines a choice between these rival cosmogonies. It clearly does

not. Everyone in the case in hand agreed that the measuring techniques were uncontroversial; everyone agreed that Descartes's cosmogony required an earth that did not bulge at the equator and that Newtonian cosmogony required an oblately spheroidal earth.

Had scientists been prepared to make Quine-like maneuvers, abandoning (say) *modus ponens*, they obviously could have held on to Cartesian physics "come what may." But that is beside the point, for if one suspends the rules of inference, then there are obviously no inferences to be made. What those who hold that underdetermination undermines methodology must show is that methodological rules, even when scrupulously adhered to, fail to sustain the drawing of any clear preferences. As this historical case makes clear, the rule cited and the relevant evidence required a choice in favor of Newtonian mechanics.

Let me not be misunderstood. I am not claiming that Newtonian mechanics was "proved" by the experiments of the Académie des Sciences, still less that Cartesian mechanics was "refuted" by those experiments. Nor would I suggest for a moment that the rule in question (R_1) excluded all possible rivals to Newtonian mechanics. What is being claimed, rather, is that this case involves a certain plausible rule of theory preference that, when applied to a specific body of evidence and a specific theory choice situation, yielded (in conjunction with familiar rules of deductive logic and of evidential assessment) *unambiguous* advice to the effect that one theory of the pair under consideration should be *rejected*. That complex of rules and evidence *determined* the choice between the two systems of mechanics, for anyone who accepted the rule(s) in question.

■ | Underdetermination and the "Sociologizing of Epistemology"

If (as we saw in the first section) some scholars have been too quick in drawing ampliative morals from QUD, others have seen in such Duhem-Quine-style underdetermination a rationale for the claim that science is, at least in large measure, the result of social processes of "negotiation" and the pursuit of personal interest and prestige. Specifically, writers like Hesse and Bloor have argued that, because theories are deductively underdetermined (HUD), it is reasonable to expect that the adoption by scientists of various ampliative criteria of theory evaluation is the result of various social, "extra-scientific" forces acting on them. Such arguments are as misleading as they are commonplace.⁴

The most serious mistake they make is that of supposing that any of the normative forms of underdetermination (whether deductive or ampliative, weak or strong) entails anything whatever about what *causes* scientists to adopt the theories or the ampliative rules that they do. Consider,

for instance, Hesse's treatment of underdetermination in her recent *Revolutions and Reconstructions in the Philosophy of Science*. She there argues that, since Quine has shown that theories are deductively underdetermined by the data, it follows that theory choice must be based, at least in part, on certain "non-logical," "extra-empirical" criteria for what counts as a good theory.¹⁰ Quine himself would probably agree with that much. But Hesse then goes on to say that:

it is only a short step from this philosophy of science to the suggestion that adoption of such [non-logical, extra-empirical] criteria, that can be seen to be different for different groups and at different periods, should be explicable by social rather than logical factors.

The thesis being propounded by these writers is that since the rules of deductive logic by themselves underdetermine theory choice, it is only natural to believe that the choice of ampliative criteria of theory evaluation (with which a scientist supplements the rules of deductive logic) are to be explained by "social rather than logical factors." It is not very clear from Hesse's discussion precisely what counts as a "social factor"; but she evidently seems to think—for her argument presupposes—that everything is either deductive logic or sociology. To the extent that a scientist's beliefs go beyond what is deductively justified, Hesse seems to insist, to that degree is it an artifact of the scientist's social environment. (Once again, we find ourselves running up against the belief—against which Duhem inveighs in the opening quotation—that formal logic exhausts the realm of the "rational.")

Hesse's contrast, of course, is doubly bogus. On the one side, it presupposes that there is nothing social about the laws of logic. But since those laws are formulated in a language made by humans and are themselves human artifacts fashioned to enable us to find our way around the world, one could hold that the laws of logic are at least in part the result of social factors. But if one holds, with Hesse, that the laws of formal logic are not the result of social factors, then what possible grounds can one have for holding that the practices that constitute ampliative logic or methodology are apt to be primarily sociological in character?

What Hesse wants to do, of course, is to use the fact of logical underdetermination (HUD) as an argument for taking a sociological approach to explaining the growth of scientific knowledge. There may or may not be good arguments for such an approach. But, as I have been at some pains to show in this essay, the underdetermination of theory choice by deductive logic is not among them.

There is another striking feature of her treatment of these issues. I refer to the fact that Hesse thinks that a semantic thesis about the relations between sets of propositions (and such is the character of the thesis of deductive underdetermination) might sustain any causal claim whatever

about the factors that lead scientists to adopt the theoretical beliefs they do. Surely, whatever the causes of a scientist's acceptance of a particular (ampliative) criterion of theory evaluation may be (whether sociological or otherwise), the thesis of deductive underdetermination entails nothing whatever about the character of those causes. The Duhem-Quine thesis is, in all of its many versions, a thesis about the logical relations between certain statements; it is not about, nor does it directly entail anything about, the causal interconnections going on in the heads of scientists who believe those statements. Short of a proof that the causal linkages between propositional attitudes mirror the formal logical relations between propositions, theses about logical underdetermination and about causal underdetermination would appear to be wholly distinct from one another. Whether theories are deductively determined by the data, or radically underdetermined by that data; in neither case does *anything* follow concerning the contingent processes whereby scientists are caused to utilize extralogical criteria for theory evaluation.

The point is that normative matters of logic and methodology need to be sharply distinguished from empirical questions about the causes of scientific belief. None of the various forms of normative underdetermination that we have discussed in this essay entails anything whatever about the causal factors responsible for scientists adopting the beliefs that they do. Confusion of the idiom of good reasons and the idiom of causal production of beliefs can only make our task of understanding either of them more difficult.⁵⁶ And there is certainly no good reason to think (with Hesse and Bloor) that, because theories are deductively underdetermined, the adoption by scientists of ampliative criteria 'should be explicable by social rather than logical factors.' It may be true, of course, that a sociological account can be given for why scientists believe what they do; but the viability of that program has nothing to do with normative underdetermination. The slide from normative to causal underdetermination is every bit as egregious as the slide (discussed earlier) from deductive to ampliative underdetermination. The wonder is that some authors (e.g., Hesse) make the one mistake as readily as the other.

David Bloor, a follower of Hesse in these matters, produces an interesting variant on the argument from underdetermination. He correctly notes two facts about the history of science: sometimes a group of scientists changes its "system of belief," even though there is "no change whatsoever in their evidential basis."⁵⁷ "Conversely," says Bloor, "systems of belief can be and have been held stable in the face of rapidly changing and highly problematic inputs from experience."⁵⁸ Both claims are surely right; scientists do not necessarily require new evidence to change their theoretical commitments, nor does new evidence—even *prima facie* refuting evidence—always cause them to change their theories. But the conclusion that Bloor draws from these two commonplaces about belief change and belief maintenance in science comes as quite a surprise. For he thinks

these facts show that *reasonable scientists are free to believe what they like, independently of the evidence*. Just as Quine had earlier asserted that scientists can hold any doctrine immune from refutation or, alternatively, they can abandon any deeply entrenched belief, so does Bloor hold that there is virtually no connection between beliefs and evidence. He writes: "So [sic] the stability of a system of belief [including science] is the prerogative of its users."⁹⁹ Here would seem to be underdetermination with a vengeance! But once the confident rhetoric is stripped away, this emerges—like the parallel Quinean holism on which it is modeled—as a clumsy non sequitur. The fact that scientists sometimes give up a theory in the absence of anomalies to it, or sometimes hold on to a theory in the face of *prima facie* anomalies for it, provides no license whatever for the claim that scientists can rationally hold on to any system of belief they like, just so long as they choose to do so.

Why do I say that Bloor's examples about scientific belief fail to sustain the general morals he draws from them? Quite simply because his argument confuses necessary with sufficient conditions. Let us accept without challenge the desiderata Bloor invokes: scientists sometimes change their mind in the absence of evidence that would seem to force them to, and scientists sometimes hang on to theories even when those theories are confronted by (what might appear to be) disquieting new evidence. What the first case shows, and all that it shows, is that the theoretical preferences of scientists are influenced by factors other than purely empirical ones. But that can scarcely come as a surprise to anyone. For instance, even the most ardent empiricists grant that considerations of simplicity, economy and coherence play a role in theory appraisal. Hence, a scientist who changes his mind in the absence of new evidence *may* simply be guided in his preferences by those of his standards that concern the nonempirical features of theory. Bloor's second case shows that new evidence is not necessarily sufficient to cause scientists to change their minds even when that evidence is *prima facie* damaging to their beliefs. Well, to a generation of philosophers of science raised to believe that theories proceed in a sea of anomalies, this is not exactly news either.

What is novel is Bloor's suggestion that one can derive from the conjunction of these home truths the thesis that scientists—quite independent of the evidence—can reasonably decide when to change their beliefs and when not to, irrespective of what they are coming to learn about the world. But note where the argument goes astray: it claims that because certain types of evidence are neither necessary nor sufficient to occasion changes of belief, it follows that no evidence can ever compel a rational scientist to change his beliefs. This is exactly akin to saying that, because surgery is not always necessary to cure gall stones, nor always sufficient to cure them, it follows that surgery is never the appropriate treatment of choice for gall stones. In the same way, Bloor argues that because beliefs sometimes change reasonably in the absence of new evidence and sometimes

do not change in the face of new evidence, it follows that we are always rationally free to let our social interests shape our beliefs.

■ | Conclusion

We can draw together the strands of this essay by stating a range of conclusions that seem to flow from the analysis:

- The fact that a theory is deductively underdetermined (relative to certain evidence) does not warrant the claim that it is ampliatively underdetermined (relative to the same evidence).
- Even if we can show in principle the nonuniqueness of a certain theory with respect to certain rules and evidence (i.e., even if theory choice is weakly underdetermined by those rules), it does not follow that that theory cannot be rationally judged to be better than its extant rivals (viz., that the choice is strongly underdetermined).
- The *normative* underdetermination of a theory (given certain rules and evidence) does not entail that a scientist's belief in that theory is causally underdetermined by the same rules and evidence, and vice versa.
- The fact that *certain* ampliative rules or standards (e.g., simplicity) may strongly underdetermine theory choice does not warrant the blanket (Quinean/Kuhnian) claim that all rules similarly underdetermine theory choice.

None of this involves a denial (a) that theory choice is always deductively underdetermined (HUD) or (b) that the nonuniqueness thesis may be correct. But one may grant all that and still conclude from the foregoing that no one has yet shown that established forms of underdetermination do anything to undermine scientific methodology as a venture, in either its normative or its descriptive aspect. The relativist critique of epistemology and methodology, insofar as it is based on arguments from underdetermination, has produced much heat but no light whatever.

■ | Appendix

In the main body of the paper, I have (for ease of exposition) ignored the more *holistic* features of Quine's treatment of underdetermination. Thus, I have spoken about single theories (a) having confirming instances, (b) entailing observation statements, and (c) enjoying given degrees of evidential support. Most of Quine's self-styled advocates engage in similar

simplifications. Quine himself, however, at least in most of his moods, denies that single theories exhibit (a), (b), or (c). It is, on his view, only whole systems of theories that link up to experience. So if this critique of Quine's treatment of underdetermination is to have the force required, I need to recast it so that a thoroughgoing holist can see its force.

The reformulation of my argument in holistic terms could proceed along the following lines. The nested or systemic version of the non-uniqueness thesis would insist that: For any theory, *T*, embedded in a system, *S*, and any body of evidence, *e*, there will be at least one other system, *S'* (containing a rival to *T*), such that *S'* is as well supported by *e* as *S* is. The stronger, nested egalitarian thesis would read: For any theory, *T*, embedded in a system, *S*, and any body of evidence, *e*, there will be systems, *S*₁, *S*₂, . . . , *S*_{*n*}, each containing a different rival to *T*, such that each is as well supported by *e* as *S*.

Both these doctrines suffer from the defects already noted afflicting their nonholistic counterparts. Specifically, Quine has not shown that, for any arbitrarily selected rival theories, *T*₁ and *T*₂, there are respective nestings for them, *S*₁ and *S*₂, that will enjoy equivalent degrees of empirical support. Quine can, with some degree of plausibility, claim that it will be possible to find systemic embeddings for *T*₁ and *T*₂ such that *S*₁ and *S*₂ will be logically compatible with all the relevant evidence. And it is even remotely possible, I suppose, that he could show that there were nestings for *T*₁ and *T*₂ such that *S*₁ and *S*₂ respectively entailed all the relevant evidence. But as we have seen, such a claim is a far cry from establishing that *S*₁ and *S*₂ exhibit equal degrees of empirical support. Thus, Quine's epistemic egalitarianism is as suspect in its holistic versions as in its atomistic counterpart.

■ | Notes

1. Pierre Duhem, *Aim and Structure of Physical Theory*, 217.
2. Lakatos once put the point this way:

A brilliant school of scholars (backed by a rich society to finance a few well-planned tests) might succeed in pushing any fantastic programme [however "absurd"] ahead, or, alternatively, if so inclined, in overthrowing any arbitrarily chosen pillar of "established knowledge." (In I. Lakatos and A. Musgrave, eds., *Criticism and the Growth of Knowledge* (Cambridge: Cambridge University Press, 1970), 187–88.)

3. See especially R. Boyd, "Realism, Underdetermination, and a Causal Theory of Evidence," *Noûs*, 7 (1973): 1–12. W. Newton-Smith goes so far as to entertain (if later to reject) the hypothesis that "given that there can be cases of the underdetermination of theory by data, realism . . . has to be rejected." ("The Underdetermination of Theories by Data," in N. R. Hilpinen, ed., *Rationality in Science* [Dordrecht: Reidel, 1980], 105.) (Compare John Worrall, "Scientific Realism and Scientific Change," *The Philosophical Quarterly*, 32 [1982]: 210–31.)

4. See chap. 2 of Hesse's *Revolutions and Reconstructions in the Philosophy of Science* (Notre Dame: Notre Dame Press, 1980) and D. Bloor's *Knowledge and Social Imagery* (London: Routledge, 1976) and "The Strengths [sic] of the Strong Programme," *Philosophy of the Social Sciences*, 11 (1981): 199-214.

5. See H. Collins's essays in the special number of *Social Studies of Science*, 11 (1981). Among Collins's many fatuous *obiter dicta*, my favorites are these:

- "the natural world in no way constrains what is believed to be" (*ibid.*, 54); and
- "the natural world has a small or nonexistent role in the construction of scientific knowledge" (3).

Collins's capacity for hyperbole is equaled only by his tolerance for inconsistency, since (as I have shown in "Collins's Blend of Relativism and Empiricism," *Social Studies of Science*, 12 [1982], 131-33) he attempts to argue for these conclusions by the use of empirical evidence! Lest it be supposed that Collins's position is idiosyncratic, bear in mind that the self-styled "arch-rationalist," Imre Lakatos, could also write in a similar vein, apropos of underdetermination, that:

The direction of science is determined primarily by human creative imagination and not by the universe of facts which surrounds us. Creative imagination is likely to find corroborating novel evidence even for the most "absurd" programme, if the search has sufficient drive (Lakatos, *Philosophical Papers* [Cambridge: Cambridge University Press, 1978], vol. 1, 99.)

6. For a discussion of many of the relevant literary texts, see A. Nehamas, "The Postulated Author," *Critical Inquiry*, 8 (1981): 133-49.

7. See, for instance, my "The Pseudo-Science of Science?," *Philosophy of the Social Sciences*, 11 (1981): 173-98; "More on Bloor," *Philosophy of the Social Sciences*, 12 (1982): 71-74; "Kuhn's Critique of Methodology," in J. Pitt, ed., *Change and Progress in Modern Science* (Dordrecht: Reidel, 1985), 283-300; "Explaining the Success of Science: Beyond Epistemic Realism and Relativism," G. Gutting et al., eds., *Science and Reality: Recent Work in the Philosophy of Science* (Notre Dame: Notre Dame Press, 1984), 83-105; "Are All Theories Equally Good?" in R. Nola ed., *Relativism and Realism in Science* (Dordrecht: Reidel, 1988), 117-39; "Cognitive Relativism," in R. Egidi, ed., *La Svolta Relativistica* (Rome: Franco Angeli, 1988) 203-24; "Relativism, Naturalism and Reticulation," *Synthese*, 71 (1987), 114-39; "Methodology's Prospects," in A. Fine and P. Machamer, eds., *PSA 1986*, vol. 2, (East Lansing, Mich.: Philosophy of Science Association) [347-54]; and "For Method: Or Against Feyerabend," in [J. R. Brown and J. Mittelstrass, eds., *An Intimate Relation: Studies in the History and Philosophy of Science: Presented to Robert E. Butts on his Sixtieth Birthday* (Dordrecht, Netherlands: Kluwer, 1989), 299-317].

8. There are, of course, more interesting conceptions of "theory" than this minimal one; but I do not want to beg any questions by imposing a foreign conception of theory on those authors whose work I shall be discussing.

9. Quine has voiced a preference that the view I am attributing to him should be called "the holist thesis," rather than a "thesis of underdetermination." (See especially his "On Empirically Equivalent Systems of the World," *Erkenntnis*, 9

(1975: 313–28.) I am reluctant to accept his terminological recommendation here, both because Quine's holism is often (and rightly) seen as belonging to the family of underdetermination arguments, and because it has become customary to use the term underdetermination to refer to Quine's holist position. I shall be preserving the spirit of Quine's recommendation, however, by insisting that we distinguish between what I call "nonuniqueness" (which is very close to what Quine himself calls "underdetermination") and egalitarianism (which represents one version of Quinean holism). (For a definition of these terms, see below [in text].)

10. It is important to be clear that Quine's nonuniqueness thesis is *not* simply a restatement of HUD, despite certain surface similarities. HUD is entirely a *logico-semantic* thesis about deductive relationships; it says nothing whatever about issues of empirical support. The nonuniqueness thesis, by contrast, is an *epistemic* thesis.

11. Obviously, the egalitarian thesis entails the nonuniqueness thesis, but not conversely.

12. Quine specifically put it this way: "Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system [of belief]." ("Two Dogmas of Empiricism," in S. Harding, ed., *Can Theories Be Refuted?* [Dordrecht: Reidel, 1976], 60 [296–97]. I am quoting from the version of Quine's paper in the Harding volume since I will be citing a number of other works included there.)

13. Grünbaum, in his *Philosophical Problems of Space and Time*, 2d ed. (Dordrecht: Reidel 1974), 590–610, has pointed to a number of much more sophisticated, but equally trivial, ways of reconciling an apparently refuted theory with recalcitrant evidence.

14. Quine in Harding (see note 12 above), 60 [297]. In a much later, backtracking essay ("On Empirically Equivalent Systems of the World," *Erkenntnis*, 9 (1975): 313–28), Quine seeks to distance himself from the proposal, implied in "Two Dogmas . . .," that it is always (rationally) possible to reject 'observation reports'. Specifically, he says that QUD "would be wrong if understood as imposing an equal status on all the statements in a scientific theory and thus denying the strong presumption in favor of the observation statements. It is this [latter] bias which makes science empirical" (*ibid.*, p. 314).

15. In fact, of course, Quine thinks that we generally do (should?) not use such stratagems. But his only argument for avoiding such tricks, at least in "Two Dogmas of Empiricism," is that they make our theories more complex and our belief systems less efficient. On Quine's view, neither of those considerations carries any epistemic freight.

16. Quine, in Harding (see note 12 above), 63. I am not alone in finding Quine's notion of pragmatic rationality to be epistemically sterile. Lakatos, for instance, remarks of Quine's "pragmatic rationality": "I find it irrational to call this 'rational'" (Lakatos, *Philosophical Papers* [Cambridge: Cambridge University Press, 1978], vol. 1, 97n).

17. W. Quine, *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969), 79.

18. See especially A. Grünbaum, *Philosophical Problems of Space and Time* (Dordrecht: Reidel 1974), 585–92 and Larry Laudan, "Grünbaum on 'the Duhemian Argument'," in S. Harding, *Can Theories Be Refuted?* (Dordrecht: Reidel, 1975).

19. In a letter to Grünbaum, published in Harding (see note 12 above), p. 132, Quine granted that "the Duhem-Quine thesis" (a key part of Quine's holism and thus of QUD) "is untenable if taken nontrivially." Quine even goes so far as to say that the thesis is not "an interesting thesis as such." He claims that all he used it for was to motivate his claim that meaning comes in large units, rather than sentence-by-sentence. But just to the extent that Quine's QUD is untenable on any nontrivial reading, then so is his epistemic claim that any theory can rationally be held true come what may. Interestingly, as late as 1975, and despite his concession that the D-Q thesis is untenable in its nontrivial version, Quine was still defending his holistic account of theory testing (see below in text).

20. And if they did not, the web would itself be highly suspect on other epistemic grounds.

21. Or, more strictly, that there is a network of statements that includes the flat-earth hypothesis and that is as well confirmed as any network of statements including the oblate-spheroid hypothesis.

22. Since I have already discussed Quine's views on these matters, and will treat Kuhn's in the next section, I will limit my illustration here to a brief treatment of Hesse's extrapolations from the underdetermination thesis. The example comes from Mary Hesse's recent discussion of underdetermination in her *Revolutions and Reconstructions in the Philosophy of Science*. She writes:

Quine points out that scientific theories are never logically determined by data, and that there are consequently [sic] always in principle alternative theories that fit the data more or less adequately. (See note 4 above, 32–33)

Hesse appears to be arguing that, because theories are deductively underdetermined, it follows that numerous theories will always fit the data "more or less adequately." But this conclusion follows not at all from Quine's arguments, since the notion of "adequacy of fit" between a theory and the data is an epistemic and methodological notion, not a logical or syntactic one. I take it that the claim that a theory fits a given body of data "more or less adequately" is meant to be, among other things, an indication that the data lend a certain degree of support to the theory that they "fit." As we have already seen, there may be numerous rival theories that fit the data (say in the sense of entailing them); yet that implies nothing about equivalent degrees of support enjoyed by those rival theories. It would do so only if we subscribed to some theory of evidential support that held that "fitting the data" was merely a matter of entailing it, or approximately entailing it (assuming counterfactually that this latter expression is coherent). Indeed, it is generally true that no available theories exactly entail the available data; so sophisticated inductive-statistical theories must be brought to bear to determine which fits the data best. We have seen that Quine's discussion of underdetermination leaves altogether open the question whether there are always multiple theories that "fit the data" equally well, when that phrase is acknowledged as having extra-syntactic import. If one is to establish that numerous alternative theories "fit the data more or less adequately," then one must give arguments for such am-

pliative underdetermination that goes well beyond HUD and any plausible version of QUD.

23. I remind the reader again that neither Quine nor anyone else has successfully established the cogency of the entailment version of QUD, let alone the explanatory or empirical support versions thereof.

24. If it did, then we should have to say that patently nonempirical hypotheses like "The Absolute is pure becoming" had substantial evidence in their favor.

25. In his initial formulation of the qualitative theory of confirmation, Hempel toyed with the idea of running together the entailment relation and the evidential relation; but he went on firmly to reject it, not least for the numerous paradoxes it exhibits.

26. Consider, for sake of simplicity, the case where two theories each entail a true evidence statement, *e*. The posterior probability of each theory is a function of the ratio of the prior probability of the theory to the prior probability of *e*. Hence if the two theories began with different priors, they must end up with different posterior probabilities, even though supported by precisely the same evidence.

27. It is generally curious that Quine, who has had such a decisive impact on contemporary epistemology, scarcely ever—in "Two Dogmas . . ." or elsewhere—discussed the rules of ampliative inference. So far as I can see, Quine generally believed that ampliative inference consisted wholly of hypothetico-deduction and a simplicity postulate!

28. As we shall eventually see, the kind of underdetermination advocated in *Word and Object* has no bearing whatever on (2*) or QUD.

29. W. V. Quine, *Word and Object* (Cambridge, Mass.: M.I.T. Press, 1960), 22, my italics.

30. *Ibid.*, p. 22, my italics. There is, of course, this difference between these two passages: The first says that commonsense talk of objects may conceivably underdetermine theory preferences, whereas the second passage is arguing for the probability that sensations underdetermine theory choice. In neither case does Quine give us an argument.

31. *Ibid.*, 23, my italics.

32. Except a vague version of the principle of simplicity.

33. *Ibid.*, 21.

34. *Ibid.*, 22–23.

35. In some of Quine's more recent writings (see especially his "On Empirically Equivalent Systems of the World," *Erkenntnis*, 9 (1975): 313–28), he has tended to soften the force of underdetermination in a variety of ways. As he now puts it, "The more closely we examine the thesis [of underdetermination], the less we seem to be able to claim for it as a 'theoretical thesis'" (*ibid.*, 326).

He does, however, still want to insist that "it retains significance in terms of what is practically feasible" (*ibid.*). Roughly speaking, Quine's distinction between theoretical and practical underdetermination corresponds to the situations we would be in if we had all the available evidence (theoretical underdetermination) and if we had only the sort of evidence we now possess (practical underdetermi-

nation). If the considerations that I have offered earlier are right, the thesis of practical Quinean underdetermination is as precarious as the thesis of theoretical underdetermination.

36. Quine does not repudiate the egalitarian thesis in *Word and Object*; it simply does not figure here.

37. In some of Quine's later gyrations (esp. his "On Empirically Equivalent Systems of the World") he appears to waver about the soundness of the nonuniqueness thesis, saying that he does not know whether it is true. However, he still holds on there to the egalitarian thesis, maintaining that it is "plausible" and "less beset with obscurities" than HUD (*ibid.*, 313). He even seems to think that nonuniqueness depends argumentatively on the egalitarian thesis, or at least, as he puts it, that the "holism thesis [egalitarianism] lends credence to the underdetermination theses [nonuniqueness]." (*ibid.*) This is rather like saying that the hypothesis that there are fairies at the bottom of my garden lends credence to the hypothesis that something is eating my carrots.

38. E.g., the difference between Quine's (0) and (1).

39. Quine's repeated failures to turn any of his assertions about normative underdetermination into plausible arguments may explain why, since the mid-1970s, he has been distancing himself from virtually all the strong readings of his early writings on this topic. Thus, in his 1975 paper on the topic, he offers what he calls "my latest tempered version" of the thesis of underdetermination. It amounts to a variant of nonuniqueness thesis. ("The thesis of underdetermination . . . asserts that our system of the world is bound to have empirically equivalent alternatives . . ." *ibid.*, 327.) Significantly, Quine is now not even sure whether he believes this thesis: "This, for me, is [now] an open question" (*ibid.*).

40. What follows is a condensation of a much longer argument, which can be found, with appropriate documentation, in my "Kuhn's Critique of Methodology" (see note 7 above).

41. Apropos the resistance to the introduction of a new paradigm, Kuhn claims that the historian "will not find a point at which resistance becomes illogical or unscientific" (*The Structure of Scientific Revolutions*, Chicago: University of Chicago Press, 1962, 159).

42. Kuhn, *The Essential Tension* (Chicago: University of Chicago Press, 1970), 325 [106]. My italics.

43. *Ibid.* [106] My italics.

44. *Ibid.*, 329 [109]. My italics.

45. In *Structure of Scientific Revolutions*, Kuhn had maintained that the refusal to accept a theory or paradigm "is not a violation of scientific standards" (159).

46. Kuhn, *The Essential Tension*, 322 [103].

47. *Ibid.*, 322 [103].

48. See, for instance, Derek Price, "Contra-Copernicus," in M. Clagett, ed., *Critical Problems in the History of Science* (Madison, 1959), 197-218.

49. A similar remark can be made about several of Popper's rules about theory choice. Thus, Miller and Tichý have shown that Popper's rule "accept the theory

with greater verisimilitude" underdetermines choice between incomplete theories; and Grünbaum has shown that Popper's rule "prefer the theory with a higher degree of falsifiability" underdetermines choice between mutually incompatible theories. [David Miller, "Popper's Qualitative Theory of Verisimilitude," *British Journal for the Philosophy of Science* 25 (1974): 166-77; Pavel Tichý, "On Popper's Definitions of Verisimilitude," *British Journal for the Philosophy of Science* 25 (1974): 155-60; Adolf Grünbaum, "Is the Method of Bold Conjectures and Attempted Refutations Justifiably the Method of Science?" *British Journal for the Philosophy of Science* 27 (1976): 105-36.]

50. Recall Quine's claim that we can hang on to any statement we like by changing the meaning of its terms.

51. See, for instance, I. Todhunter, *History of the Theories of Attraction and the Figure of the Earth* (New York: Dover, 1962).

52. Typically, astronomical measurements of angles subtended at meridian by stipulated stars were used to determine geodetic distances.

53. In fact, the actual choice during the 1730s, when these measurements were carried out, was between a Cassini-emended version of Cartesian cosmogony (which predicted an oblong form for the earth) and Newtonian cosmology (which required an oblate shape).

54. Indeed, most of so-called radical sociology of knowledge rests on just such confusions about what does and does not follow from underdetermination.

55. M. Hesse (see note 4 above), 33.

56. This is not to say, of course, that there are no contexts in which it is reasonable to speak of reasons as causes of beliefs and actions. But it is to stress that logical relations among statements cannot unproblematically be read off as causal linkages between propositional attitudes.

57. Bloor, "Reply to Buchdahl," *Studies in History and Philosophy of Science*, 13 (1982): 306.

58. *Ibid.*

59. *Ibid.* In his milder moments, Bloor attempts to play down the radicalness of his position by suggesting (in my language) that it is the nonuniqueness version of underdetermination rather than the egalitarian version that he is committed to. Thus, he says at one point that "I am not saying that any alleged law would work in any circumstances" ("Durkheim and Mauss Revisited," *Studies in History and Philosophy of Science*, 13 [1982]: 273). But if indeed Bloor believes that the stability of a system of belief is the prerogative of its users, then it seems he must hold that any "alleged law" could be made to work in any conceivable circumstances; otherwise, there would be some systems of belief that it was not at the prerogative of the holder to decide whether to hang on to.

- 3.1 Duhem's Holism and the Ambiguity of Falsification** 355
The Role of Auxiliary Hypotheses in the Experimental Testing of Theories 355
Duhem and the Theory-Ladenness of Observation 358
Why Crucial Experiments Are Impossible in Physics 359
Duhem's Critique of Inductivism: The Attack on Newtonian Method 363
Duhem's Rejection of Poincaré's Conventionalism 364
- 3.2 Quine's Attack on the Two Dogmas of Empiricism** 365
The Two Dogmas: Analyticity and Reductionism 366
The Status of Mathematical Truths 367
Quine's Rejection of the Analytic-Synthetic Distinction 372
Quine's Holism and His Rejection of A Priori Knowledge 373
- 3.3 Gillies on Duhem and Quine** 374
Three Contrasts between the Duhem Thesis and the Quine Thesis 374
Gillies' Criticisms of Quine 375
Gillies' Criticisms of Duhem 376
Gillies' Version of the Duhem-Quine Thesis 381
- 3.4 Laudan's Repudiation of Underdetermination** 382
Humean Underdetermination (HUD) 383
Quinean Underdetermination (QUD) 385
Goodman's Gruesome New Riddle of Induction 389
Kuhn's Thesis of Local Underdetermination 391
The Strong Programme in the Sociology of Science 393
- 3.5 Summary** 397

3 | . COMMENTARY

3.1 | Duhem's Holism and the Ambiguity of Falsification

Duhem begins "Physical Theory and Experiment" with some lengthy quotations from his fellow countryman, Claude Bernard (1813–1878). Bernard was hailed, then as now, as the father of experimental physiology. In his famous work on scientific method, *An Introduction to the Study of Experimental Medicine* (1865), he rejected the narrow inductivist conception of science that he associated (probably falsely) with writers such as Francis Bacon. The narrow inductivist sees science as a matter of simply generalizing from observation and experiment, untainted by hypotheses and theories.¹ According to the narrow inductivist, scientists should perform experiments and make observations without having any prior theory to guide them.

Bernard, quite rightly, regarded this sort of inductivism as absurd. Theories play an indispensable role in science, in part by suggesting new experiments and observations that can be used to confirm or refute those theories and thus advance our knowledge. But Bernard stressed the need for scientists to maintain an open mind and to adopt a skeptical attitude toward all theories when performing experiments (or making observations) to test a theory. When testing a theory, experimenters should leave to one side all theoretical preconceptions and beliefs in order to preserve their impartiality. Theory proposes (by suggesting experimental tests), but experiment disposes. Should experiment disagree with theoretical predictions, then the theory making those predictions ought to be rejected; should experiment agree with theoretical predictions, then the theory is confirmed. What is important is to preserve the objectivity and impartiality of observation and experiment by striving to prevent the intrusion of theoretical beliefs into the assessment of experimental results. Thus, Bernard advocates that scientists wear two hats: a theoretician's hat when outside the laboratory and an experimentalist's strictly nontheoretical hat when inside the laboratory.

THE ROLE OF AUXILIARY HYPOTHESES IN THE EXPERIMENTAL TESTING OF THEORIES

Duhem finds much to commend in Bernard's experimental method for physiology, but he thinks it is too simplistic for theory testing in physics. The crucial difference between applied sciences such as physiology and fundamental sciences such as physics is that the physiologist accepts as

established fact (given by the supposedly infallible physicist) the theories that specify how the experimental apparatus and instruments work, whereas the physicist is keenly aware that the theories underlying the use of such apparatus and instruments are no less vulnerable than the specific theory that is being tested. When the physicist performs an experiment to test a theory, it is not just that particular theory that is involved in the logic of testing, but a whole range of theories and assumptions about the working of the experimental apparatus and instruments.²

Thus Duhem is led to espouse his famous thesis of holism: when a physical theory is tested by experiment, it is not that theory alone, but a large collection of theories, auxiliary hypotheses, and assumptions that are being put to the test. Let the theory being tested be T , the auxiliary hypotheses and assumptions be A_1, A_2, \dots, A_n , and the testable prediction be O_1 . Duhem is claiming that T alone will not yield the prediction O_1 ; to deduce O_1 requires not only T but also A_1, A_2, \dots, A_n . Schematically, Duhem's holist thesis is.

$$D1 \quad \sim (T \rightarrow O_1), \text{ and}$$

$$D2 \quad (T \ \& \ A_1 \ \& \ A_2 \ \& \ \dots \ \& \ A_n) \rightarrow O_1.$$

From (D1) and (D2) it is but a short step to Duhem's central claim about the ambiguity of falsification. Suppose we perform an experiment and find that O_1 is false. In other words, we discover that $\sim O_1$. Because T alone does not entail O_1 , we cannot conclude that T is false. All that follows, logically, is that at least one of T, A_1, A_2, \dots, A_n is false, and logic alone will not tell us where to pin the blame. Schematically, Duhem's minimal thesis about the ambiguity of falsification is:

$$D3 \quad \sim (\sim O_1 \rightarrow \sim T), \text{ and}$$

$$D4 \quad \sim O_1 \rightarrow \sim (T \ \& \ A_1 \ \& \ A_2 \ \& \ \dots \ \& \ A_n)$$

Although Duhem's minimal thesis about the ambiguity of falsification (D3 and D4) is important, it is not, by itself, very exciting. It is important because it contradicts many oversimplified accounts of theory testing in science. And it is certainly not trivial.³ But it does not directly imply anything about the impossibility of falsifying a theory, since it may be possible for scientists to establish that the entire set of A_1, A_2, \dots, A_n is true, thus pointing the arrow of *modus tollens* squarely at T . Even less does Duhem's minimal thesis imply anything about the possibility, in general, of always being able to retain belief in a theory no matter what the contrary evidence, simply by changing the auxiliary hypotheses. To draw wider implications such as these from Duhem's minimal thesis requires further

assumptions, many of which are quite controversial. (Which, if any, of these wider implications were endorsed by Duhem himself is also unclear.) Further claims about falsification (some by Duhem, some merely influenced by Duhem) are addressed in the later sections of this commentary. For the moment let us focus on Duhem's holism and his minimal claim about falsification.

The real issue is Duhem's holism, summarized in (D1) and (D2). For, if (D1) and (D2) are true, then (D3) and (D4)—Duhem's minimal thesis—follow as an immediate logical consequence. So, why should we agree with Duhem that theories in physics entail testable consequences only with the help of numerous auxiliary hypotheses, theories, and assumptions? Consider a particular theory, say Newton's theory of universal gravitation. According to Newton's gravitational theory, every piece of matter attracts every other piece of matter with a force that varies directly with the product of the masses involved and inversely with the square of their distance apart. The law can be summarized in the familiar form:

$$F_{12} = G \frac{m_1 m_2}{d^2}.$$

Clearly, Newton's gravitational theory does not logically imply any testable prediction about the position or velocity of a particular body, such as a planet, at a particular time. For that we need more information and some additional theories and hypotheses. For example, we need Newton's three laws of motion; we need to know the initial conditions, such as the mass of the sun, the position and velocity of the planet, and its distance from the sun at a particular time; and we need to assume that no other significant masses will affect the planet gravitationally or by any other means, such as by collision. Moreover, the velocity of the planet is not something we can determine simply by looking. For that we need to use instruments, such as telescopes, cameras, and clocks. These instruments, in turn, require us to use theories to move from what we directly observe—things such as pointer readings, telescopic images, and streaks on a photographic plate—to judgments about position and velocity at particular times. Even then, what we report as *the* value of a reading on an instrument will often be the mean of several values spread over a small range. Thus, for Newton's gravitational theory, Duhem's holist thesis (D1 and D2) is quite plausible. But showing that holism is true of one theory does not show that holism is true of all theories. For that, a general argument is needed. The argument to which Duhem and others most frequently appeal is based on the theory-ladenness of observation.

DUHEM AND THE THEORY-LADENNESS OF OBSERVATION

Duhem defends the thesis of the theory-ladenness of observation in part 2, chapter 4 of *Aim and Structure*. His version of the thesis is restricted to experiments performed to test theories in physics. In Duhem's own words:

An experiment in physics is not simply the observation of a phenomenon; it is, besides, the theoretical interpretation of this phenomenon. (144)

An experiment in physics is the precise observation of the phenomena accompanied by an *interpretation* of these phenomena; this interpretation substitutes for the concrete data really gathered by observation abstract and symbolic representations which correspond to them by virtue of the theories admitted by the observer. (147)

We can best appreciate Duhem's claim by comparing the language of theories in physics with the language of observation. Physical theories typically employ terms such as *current*, *voltage*, *force*, *pressure*, *entropy*, and *temperature* in order to formulate functional laws in the form of mathematical equations. When an untrained person enters the physicist's laboratory, he does not "see" that a current of 0.75 amps is flowing through a resistor, or that the temperature of molten titanium indicated by an optical pyrometer is 1800 degrees Celsius. Rather, he would report seeing various pieces of equipment, some connected to others with wires, some with pointers moving around dials, others with eyepieces through which various images can be seen. What the physicist can deduce from the physical theory, in conjunction with other theories, background assumptions, and initial conditions, are predictions using the theoretical language of amps, volts, degrees Celsius, and so on. In order to connect these predictions with direct observation, the physicist needs to translate from the everyday language of the untrained observer to the theoretical language of the physicist. This translation is effected by using theories about how the measuring instruments work. In Duhem's terminology we convert practical facts into theoretical facts, where practical facts, reported in the observation language of the scientifically untrained observer, are what are "really observed" as Duhem puts it (151). In this way Duhem argues that theory plays an indispensable role in physical experiments and thus, contrary to Bernard, that physicists cannot leave behind all theory when they enter the laboratory.

Thus, the theory-ladenness thesis gives us a general argument for Duhem's holist thesis: for any theory that is not expressed entirely in the language of observation, that theory will logically imply observation statements (expressed entirely in the language of observation) only with the aid of further theories about how measuring instruments and other pieces of apparatus function.

WHY CRUCIAL EXPERIMENTS ARE IMPOSSIBLE IN PHYSICS

One of the most famous aspects of Duhem's philosophy of science is his denial of the possibility of crucial experiments in physics. The term *crucial experiment* (*experimentum crucis*—literally, “experiment of the cross”) was coined by Francis Bacon in the seventeenth century. By the nineteenth century, it had come to signify an experiment that conclusively falsifies one of two rival theories or hypotheses, thus establishing its rival as well confirmed or true. It is important to bear in mind both of these characteristics: the power to refute one theory conclusively and, in virtue of this, the presumed ability to establish its rival as well confirmed or true.

Duhem gives two examples of experiments regarded as crucial by most scientists in the nineteenth century, both from optics. The first is Otto Wiener's experiment (published in 1890) on the direction of vibration of polarized light. By the middle of the nineteenth century, the wave theory of light had gained widespread acceptance, displacing the earlier particle theory of Newton, Laplace, and Biot. According to the wave theory, light is a transverse vibration in a medium, the ether, that occupies all of space and permeates material bodies. Just like waves traveling down a rope, the vibrations that constitute light are at right angles to the direction in which the wave is traveling. One of the directions perpendicular to the direction of travel is the plane of polarization. The plane of polarization is defined purely conventionally in terms of operations that can be carried out in the laboratory. The interesting theoretical question was, Do light waves vibrate parallel to the plane of polarization, or perpendicular (normal) to it? According to the theory of Augustin Fresnel, the vibration is perpendicular to the plane of polarization. According to the theory of F. E. Neumann and James MacCullagh, the vibration is parallel to it. Wiener's ingenious experiment distinguished between these two opposed predictions and verified the prediction made by Fresnel. This experimental result was widely interpreted at the time as decisively refuting the Neumann-MacCullagh theory and confirming Fresnel's theory. Indeed, some physicists regarded Wiener's experiment as showing that the vibration of light was not merely a mathematical concept or a convenient theoretical postulate but a physical reality with real, dynamical effects.

The other experiment cited by Duhem as one that nearly every nineteenth-century scientist regarded as crucial was Jean Foucault's accurate measurement of the velocity of light in air and water, which was seen as a crucial test between the wave theory of light and its rival, the particle theory. As Descartes and Newton had realized, if light consisted of tiny particles, then refraction would occur when light enters a medium denser than air because of short-range forces that briefly accelerate the particles. Consequently, the particle theory predicts that the velocity of light in water (or glass, or any medium denser than air) must be greater than the velocity of light in air. The wave theory makes precisely the opposite prediction.

Both theories entail that the ratio of velocities equals the ratio of the refractive indices, but the wave theory predicts that light travels more slowly in water than it does in air. Foucault's measurements, showing that light travels more slowly in water than in air, were heralded as the definitive refutation of the particle theory and a spectacular confirmation of the wave theory.⁴

Duhem regards neither Wiener's nor Foucault's experiment as crucial. For in both cases, what is in question is not a single proposition but a whole system of hypotheses, theories, and assumptions that are needed to derive testable predictions. Thus, the negative result obtained, for example, in Wiener's experiment shows only that at least one of these elements is false, but it does not tell us which one. Duhem's holism entails that no experiment can conclusively falsify a physical theory, so, *a fortiori*, there can be no crucial experiments in physics. This is Duhem's minimal thesis about crucial experiments: logic alone does not dictate that we have to abandon a physical theory as false when its predictions disagree with experiment. In other words, Duhem's minimal thesis about crucial experiments is simply a restatement of (D3).

Duhem contrasts the use of experiments to refute theories in physics with *reductio ad absurdum* reasoning (or the method of indirect proof) in mathematics and logic. In mathematics, many important theorems can be proven using this method. First, one assumes that the proposition one wants to prove is false. Then, one deduces from the falsity of the proposition a contradiction, using other mathematical and logical truths as additional premises. Unlike the refutation of physical theories by experiment, *reductio ad absurdum* reasoning is logically conclusive. The essential difference between the two is that when mathematicians use *reductio ad absurdum* reasoning, all the additional assumptions they adopt as premises are necessary truths. Since all these other premises are necessary truths (and thus, let us assume, known to be true with absolute certainty), when mathematicians validly deduce a contradiction from them in conjunction with the proposition that they assumed, for the sake of argument, to be false, it follows that the proposition in question must be true. But physicists deal not with necessary truths, but with empirical assumptions and hypotheses any number of which might be false. Thus, experiments in physics cannot unequivocally show that the theory being tested is false when it yields false predictions.

At the end of the third section of his article, dealing with the impossibility of crucial experiments, Duhem extends the analogy with *reductio* reasoning a little further and makes a point of considerable logical and methodological importance. Suppose, Duhem says, that experimental reasoning in physics were just like *reductio* reasoning in mathematics and that, when faced with the situation represented by (D4),

$$D4 \quad \sim O_1 \rightarrow \sim (T \& A_1 \& A_2 \& \dots \& A_n),$$

all the auxiliary assumptions, A_1 through A_n , were truths "compelled to be necessary by strict logic" (265) just as they are in mathematics. In that case, we would have a logically conclusive refutation of theory T by the observation statement $\sim O_1$. And in mathematics, that would suffice to establish that $\sim T$ is a necessary truth. But in physics, the alternative to theory T is not its logical contrary $\sim T$ but some rival theory, T^* . The truth of T^* does not follow logically from the falsity of T . Symbolically:

$$D5 \quad \sim (\sim T \rightarrow T^*).$$

In optics, for example, the wave theory and the particle theory are not the only possible theories of the nature of light. Thus even if Foucault's experiment could falsify the wave theory conclusively, it still would not follow that his experiment was crucial because a crucial experiment is traditionally defined as one that not only falsifies one theory but also thereby establishes its rival as true (or well confirmed). As Duhem says at the end of section 3, in order for the experimental method in physics to have the same conclusive power as *reductio* reasoning in mathematics, physicists would have to be able "to enumerate completely the various hypotheses which may cover a determinate group of phenomena" (266), something that lies beyond their abilities.⁵ Without such a complete enumeration, the inference to the truth of a physical theory would be no better than a toss of a coin; and, as Duhem reminds us, "The truth of a physical theory is not decided by heads or tails."

We have seen that Duhem has two reasons for denying the possibility of crucial experiments in physics. One is the familiar argument from the ambiguity of falsification, summarized in the holist thesis (D4). The other rests on the fact, summarized in (D5), that rival theories in physics are not logically exhaustive: while two rivals cannot both be true, they might both be false; thus, we cannot infer the truth of one from the falsity of the other.

Before leaving this section on crucial experiments, a few words about the significance of Duhem's position will set the scene for later discussion. Duhem took care to express his views in a qualified and guarded manner. For example, he restricted his discussion of holism to theories in physics where he thought special conditions obtained that are not found in other, less abstract, branches of science. Moreover, the target of Duhem's attack was the extreme position that experiments have the power to falsify theories conclusively, solely as a matter of deductive logic, and that, in virtue of this falsifying power, crucial experiments can establish a rival theory as true with certainty. In chastising this position as unwarranted, Duhem left open the possibility that experiments (in conjunction with other considerations) could lead rationally to the rejection of theories as false and that experiments (and more generally, successful predictions) could confirm theories to a significant degree. Only someone who (like Popper) thinks

that the logic of science is limited solely to deductive logic and who dismisses inductive inference as a myth would be seriously nonplussed by Duhem's arguments.⁶

Duhem himself was a practicing physicist who regarded many scientific theories as having been refuted partly as a result of experiments. To be sure, Duhem did not think that any theory could be shown to be false in the same conclusive, deductive manner distinctive of mathematics and logic. But, nonetheless, theories do get refuted in science, and Duhem never denied the fact. (For example, Duhem agreed that Fresnel's wave theory of light triumphed over its opposition.) It is true that in *Aim and Structure*, Duhem speculated about how the proponents of a theory might try to protect their theory from refutation by modifying one or more of the auxiliary hypotheses with which it is associated. Consider a theory T that, in conjunction with auxiliary hypotheses A_1 through A_n , predicts the observation statement O_1 . Experiment reveals that O_1 is false ($\sim O_1$). So we can no longer believe that the system of T conjoined with A_1 through A_n is true. But instead of giving up T , we replace the auxiliary hypothesis A_1 , say, with a new auxiliary, B_1 , so that $(T \ \& \ B_1 \ \& \ A_2 \ \& \ . . . \ \& \ A_n)$ is logically consistent and no longer entails $\sim O_1$. In this way, theory T could be saved from refutation. But nowhere does Duhem claim that this strategy can always be *reasonably* employed to save any theory whatever from refutation.

For example, when discussing the possibility of reconciling the particle theory of light with Foucault's experiments, Duhem writes: "If physicists had attached some value to this task, they would undoubtedly have succeeded in founding on this assumption [that light is a swarm of particles] a system of optics that would agree with Foucault's experiment" (263). The key phrase here is "if physicists had attached some value to this task." For Duhem believed that some ways of trying to save a theory in the face of experiment would be unreasonable and lack scientific value. In order for the replacement of the system $(T \ \& \ A_1 \ \& \ A_2 \ \& \ . . . \ \& \ A_n)$ by the system $(T \ \& \ B_1 \ \& \ A_2 \ \& \ . . . \ \& \ A_n)$ to be reasonable, some constraints have to be placed on the new system $(T \ \& \ B_1 \ \& \ A_2 \ \& \ . . . \ \& \ A_n)$. We have already mentioned one obvious constraint: the new system must be logically consistent. Furthermore, the new auxiliary hypothesis B_1 cannot be a proposition that is known or reasonably believed to be false. Presumably, too, the new hypothesis must not be perniciously ad hoc. The mere logical possibility of saving T from refutation by exchanging B_1 for A_1 does not imply that this will always be a reasonable move or one that has any scientific value. Indeed, as we shall see below, Duhem attacked the views of conventionalists (such as Henri Poincaré) who argued that many fundamental hypotheses of physical theory are immune to refutation.

DUHEM'S CRITIQUE OF INDUCTIVISM: THE ATTACK ON NEWTONIAN METHOD

One of the novelties of *Aim and Structure*, compared with the earlier papers on which it was based, is Duhem's attack on the naive form of inductivism espoused by Newton and others. As we have seen, Duhem rejected the claim that the experimental method can provide physicists with anything approaching absolute certainty: experiments cannot conclusively falsify theories, nor can crucial experiments conclusively prove that any theory is true. Nonetheless, some famous scientists (such as Newton and Ampère) claimed absolute certainty for their theories on the grounds that those theories were deduced directly from experimental laws. Specifically, Newton claimed that his theory of universal gravitation could be broken down into its separate postulates and each postulate could then be established as true by "deduction from the phenomena," principally Kepler's laws of planetary motion. Ampère, emulating Newton, made a similar claim for his theory of electrodynamics. (Duhem criticizes Ampère in section 5, which has been omitted from our selection.)

Now, as Duhem realized, the term *deduction* was used rather loosely prior to the twentieth century to include not just valid deductive arguments but also inductive inferences and arguments to the best explanation. For example, Newton's four Rules of Reasoning in Philosophy, which play an important role in the derivation of the law of universal gravitational attraction in the *Principia*, are rules for inductive generalization, and both Newton and Duhem were clearly aware of this fact. So Duhem does not criticize Newton and others because their arguments are not deductive. Rather, Duhem does what any responsible critic should do: he offers his criticism in light of what his opponent means by the terms he uses. Duhem aims to show that even when we understand "deduction" to include inductive generalization and the like, it still is false that each postulate of a theory such as Newton's can be inferred ("deduced") from experimental laws.

Given Duhem's holist thesis (D1 and D2), no theory and hence no postulate of a theory can, by itself, yield any testable prediction. Thus, one might expect that Duhem would criticize Newton on the grounds that the separate parts of Newton's theory cannot be linked with observation (or with experimental laws derived from observation) by any form of inference. But Duhem gives a completely different argument against Newton, an argument that has proved to be a powerful weapon against the cruder forms of inductivism in the hands of philosophers such as Karl Popper, Imre Lakatos, and Paul Feyerabend. Duhem points out that if Newton's theory of universal gravitation is true, then Kepler's laws of planetary motion are false. Because of perturbation from other planets, no planet or planetary satellite moves in a perfect ellipse, as Kepler's laws demand. Similarly, the constant in Kepler's third law (vital for deriving the inverse

square variation of gravitational attraction with distance) in fact is not constant but differs slightly for each planet, since it depends not only on the mass of the sun but also on the mass of the particular planet. Since Newton's theory and Kepler's laws are incompatible, Newton's theory cannot be derived from Kepler's laws either by deductive argument or inductive generalization.

Duhem's refutation of Newton's official account of his method is as simple as it is compelling. Duhem's own view of the relation between Newton's theory and the observed motions of the planets is holistic: only when Newton's theory is taken in its entirety, along with the laws of motion and the theories of the measuring instruments, can one logically deduce predictions about planetary motions that agree with observation.⁷

DUHEM'S REJECTION OF POINCARÉ'S CONVENTIONALISM

Although Duhem is often cited by those who insist that any theory can, in principle, be preserved in the face of recalcitrant evidence, Duhem himself denied that any theory or any theoretical postulate in physics is immune to refutation. Duhem was particularly concerned to deny the doctrine of conventionalism advocated by Henri Poincaré and others. Poincaré maintained that many of the fundamental postulates of physics such as Newton's laws of motion and the conservation principles are not empirical hypotheses at all but definitions adopted by convention. As such, no experiment or observation could ever refute them.

Both Duhem and Poincaré were writing before the revolutions in relativity theory and quantum mechanics that overthrew Newtonian mechanics and Euclidean geometry as the last word about the motions of bodies and the geometry of space. Thus, neither participant in the debate over conventionalism could appeal to history as we now can. Since Newtonian mechanics and Euclidean geometry (as a theory of physical space) have been refuted, it seems clear to us now that Duhem was right and Poincaré wrong. But at the time, Duhem could attack Poincaré only on general methodological grounds.

Duhem's position in sections 8 and 9 of his article is, as one might expect, holistic. He agrees with Poincaré that individual postulates such as Newton's first law of motion (the principle of rectilinear inertia) cannot be tested directly by any experiment. But, contrary to Poincaré, Duhem does not take this to imply that these and other laws are irrefutable or that they are really definitions in disguise. Newton's first law can be tested indirectly, as can any hypothesis in science, by being part of a system of hypotheses that makes testable predictions. In a discussion resembling that of Lakatos regarding the methodology of scientific research programmes (see "Science and Pseudoscience" in chapter 1), Duhem explains that an individual postulate can be refuted if the system to which it belongs fails to represent reality as well as some rival system of hypotheses that contains

a different postulate. When the new system replaces the old, the postulate is, in effect, refuted.

In the short final section (section 10), of his article, Duhem concludes with some brief remarks about the importance of what he calls “good sense” in science. Logic alone can never force scientists to abandon one theory or adopt another. For with sufficient ingenuity, an ailing theory can usually be repaired or modified so as to avoid falsification by experimental evidence. But, Duhem writes, after a while

it may be that we find it childish and unreasonable . . . to maintain obstinately at any cost, at the price of continual repairs and many tangled-up stays, the worm-eaten columns of a building tottering in every part, when by razing these columns it would be possible to construct a simple, elegant, and solid system. (278)

Duhem offers the work of the French physicist Jean Biot (1774–1862) as an illustration of scientific good sense in operation. For many decades in the nineteenth century, Biot (following Newton and Laplace) defended the particle theory of light. But this became increasingly difficult as the wave theory of light scored a number of successes in the hands of Thomas Young (1773–1829) and Augustin Fresnel (1788–1827). By midcentury, the vast majority of physicists had switched to the wave theory, yet Biot persisted with the particle theory, making continual modifications to the theory, often of a complex and arbitrary nature, in order to keep pace with the wave theory. But then came Foucault’s experiment showing that light traveled more slowly in water than in air. Logic could not force Biot to abandon the particle theory because, in principle, its associated auxiliary hypotheses could still be modified to agree with Foucault’s result. Foucault’s experiment was not, according to Duhem, a crucial experiment in the logical sense. But Biot’s good sense told him that enough was enough. It was time to give up the particle theory as a lost cause in favor of the much more elegant and fruitful wave theory. Thus, the particle theory was refuted by being part of a system of hypotheses that was eventually replaced by another, better system.

3.2 | Quine’s Attack on the Two Dogmas of Empiricism

Although only thirty-five years separate Duhem’s death (in 1916) from the publication of Quine’s “Two Dogmas of Empiricism” (in 1951), there is world of difference between the work of the two men. For the first half of his professional life, Duhem was a working physicist, specializing in thermodynamics. Later, he achieved renown as a historian and philosopher of science. Understandably, then, Duhem is interested primarily in

science. His theses about holism and the ambiguity of falsification concern scientific theories, usually in physics, and are supported by examples drawn from the sciences. Quine, by contrast, is a philosopher interested primarily in logic and the philosophy of language. Much of "Two Dogmas" is a reaction to the views about meaning and analyticity held by Rudolf Carnap, with whom Quine studied and interacted in the 1930s and 1940s. Quine's work is not specifically about science at all, based as it is on general considerations about language, meaning, and empiricism. But it has implications for the philosophy of science that Quine alludes to in some brief, oft-cited passages towards the end of his article. Significantly, Quine mentions Duhem only once, in a footnote that was added to his paper when it was reprinted in 1953.⁸ In "Two Dogmas" Quine was not consciously developing Duhem's line of thought or reacting to what Duhem had written. We should not, then, expect to find obvious similarities between Quine and Duhem given the differences in their interests and approaches and Quine's ignorance of Duhem's work.⁹ A detailed comparison of the views of Duhem and Quine can be found in the discussion of Donald Gillies' essay later in this commentary; Quine's doctrine of empirical underdetermination is discussed and criticized in the section on Larry Laudan's paper. In this section, our goal is to provide a helpful guide to Quine's somewhat technical and abstract paper, supplementing where necessary the discussion in Gillies' article. Since our primary concern is philosophy of science, not philosophy of language, we shall skip some of the more technical parts of Quine's paper. Our aim, here, is to sketch the shape of Quine's reasoning so as to bring out its relevance to the issues of holism, falsification, and underdetermination.

THE TWO DOGMAS: ANALYTICITY AND REDUCTIONISM

As Quine explains in his opening paragraph, the two dogmas of empiricism that he wishes to criticize are analyticity and reductionism. Analyticity usually goes by the longer title of the analytic-synthetic distinction. It is the doctrine that some statements (analytic ones) are true (or false) solely in virtue of the meanings of their constituent terms and completely independently of empirical matters of fact.¹⁰ Statements that are not analytic are said to be synthetic. Unlike their analytic brethren, synthetic statements are true (or false) depending on the way the world is. (For the sake of convenience and brevity, we shall assume that the analytic and synthetic statements under consideration are true, unless otherwise specified.)

Reductionism is the doctrine, popular among the logical positivists (and fellow travelers, such as operationalists) that every meaningful, synthetic statement is logically equivalent to some sentence containing only experiential (or observational) terms joined together with logical connectives (such as *and*, *or*, and *if . . . then . . .*). The classic phenomenalist version of reductionism (favored by empiricist philosophers from George

Berkeley in the eighteenth century to A. J. Ayer and Rudolf Carnap in the twentieth) insists that everyday statements about physical objects (e.g., "there is a yellow table in the room") are logically equivalent to statements referring solely to the experiences of one or more people. In its positivist and operationalist versions, theoretical statements in scientific theories are supposed to be equivalent to statements referring only to observable objects and operations that can be performed on them. (For example, behavioral psychologists such as John Watson and B. F. Skinner claim that all theoretical statements involving mental terms such as *belief* and *anxiety* can be reduced to statements referring only to observable behavior.) In short, reductionism claims that all synthetic statements are logically equivalent to statements in some special empiricist language—either the subjective language of immediate experience or the objective language of physical things.

Quine claims that, at bottom, the two dogmas are identical. He gives his reasons for this judgment towards the end of section 5 of his paper. First, he notes that subscribing to the verification theory of meaning entails believing that each synthetic (empirical) statement is equivalent in meaning to some set of observation statements (or, for the phenomenalist, statements about sensory events). Even those empiricists who have given up the verification theory of meaning often adopt a view of confirmation according to which each synthetic statement is uniquely associated with a set of observation statements that can either confirm or disconfirm it. (Quine uses the word *infirm* to mean *disconfirm*.) In either case, whether one is a strict, old-fashioned reductionist about meaning or what Quine regards as a closet reductionist about confirmation, one is naturally led to posit a limiting class of true statements—the analytic ones—that have no empirical meaning and that are confirmed no matter how the world is. Since these statements lack any factual component, it is tempting to conclude that their truth depends entirely on their linguistic component, on the meanings of their constituent words. In short, Quine thinks that reductionism leads to the analytic-synthetic distinction because it tempts one into positing the existence of a special class of sentences, the analytic ones, whose factual component is nil.¹¹

THE STATUS OF MATHEMATICAL TRUTHS

One can appreciate the seductive quality of the chain of reasoning that Quine describes as leading from reductionism to analyticity by considering how the members of the Vienna Circle viewed logic and mathematics. The members of the Vienna Circle were logical positivists and thus committed empiricists, verificationists, and reductionists. In the 1920s and 1930s logical positivists such as Hans Hahn, Rudolf Carnap, A. J. Ayer, and Carl Hempel were struck by the fact that we have certain knowledge of simple mathematical and logical truths that is completely independent

of experience. How is this possible? How can a priori mathematical knowledge be reconciled with empiricism (according to which all empirical knowledge depends on experience)?

One answer had been given in the eighteenth century by Immanuel Kant, who coined the terms *analytic* and *synthetic*. Kant defined a true statement as analytic if and only if the concept of the predicate is contained in the concept of the subject. (In categorical statements such as "All dogs are brown," "Some dogs are brown," and "No dogs are brown," the subject term is *dogs* and the predicate term is *brown*.) Any statement that is not analytic is synthetic. To Kant it seemed obvious that many mathematical statements, especially in geometry, were not analytic but synthetic. A typical example of a mathematical truth Kant judged to be synthetic is the proposition that a straight line is the shortest distance between two points. As far as Kant was concerned, the concept of the shortest distance between two points is not included in the concept of a line being perfectly straight. (Kant reasoned that the former is a quantitative concept, involving numerical comparisons, but the latter, he thought, is a purely qualitative idea. Hence, the former is not contained in the latter.) Contrast this with the proposition that all material bodies are extended in space, which Kant judged to be analytic because the concept of spatial extension is contained in our concept of a material body. Thus Kant concluded that strict empiricists (such as David Hume) were wrong to assume that all synthetic propositions must be a posteriori (that is, derived from experience). Mathematics (and, Kant thought, some important general principles of science, metaphysics, and ethics) are synthetic, a priori truths.

Kant's way of drawing the analytic-synthetic distinction has been widely criticized, especially by philosophers such as Gottlob Frege (1848–1925). According to Frege, the main flaws in Kant's characterization of analyticity are its restriction to categorical statements and its vague and psychological character. Many statements have a complex internal structure that resists analysis in subject-predicate form. An example is Euclid's fifth postulate: through any point outside of a given straight line there is one and only one straight line that is parallel to that given line. Another example is the following complex but necessarily true proposition: if no person has more than N hairs on his or her head and if the population of a city is greater than N , then at least two people in that city must have exactly the same number of hairs on their heads. Since Kant's definition of analyticity in terms of the predicate being contained in the subject is applicable only to categorical statements, it cannot give a complete account of analyticity. But, Frege argues, Kant's account is defective even for categorical statements because it is vague and psychological in character. When we consider whether the subject of a categorical statement already contains its predicate, we are engaging in a form of psychological introspection. This procedure is too subjective and variable to serve as an

adequate account of what makes a statement analytic. For, being subjective, it is possible that two people might use the same procedure and yet arrive at conflicting judgments about the status of a particular proposition.

Frege's alternative to Kant's proposal is that a proposition is analytic if and only if it can be proved (or disproved) from definitions using only the laws of logic. In other words, a statement is analytic if it is a tautology or can be reduced to a tautology by using definitions to substitute some terms for others meaning the same thing. Thus, for example, "all bachelors are bachelors" is a tautology and hence analytic; "all bachelors are unmarried" is analytic, since it reduces to a tautology when "unmarried male" is substituted for "bachelor." In contrast with Kant, Frege espoused the doctrine of logicism, the thesis that a significant part of mathematics, namely arithmetic, is ultimately reducible to logic. We have a priori knowledge of arithmetic because arithmetical truths are analytic; arithmetical truths are analytic because they reduce to logical truths. They are known to be true independently of experience because, by means of definitions of concepts such as *number* and *is the successor of*, they reduce to the laws of logic and we see immediately, by the use of our reason, that these laws of logic must be true. (Concerning geometry, Frege accepted Kant's view that geometrical truths are synthetic a priori.) But why do we have a priori knowledge of logic? At this point Frege was an unabashed Platonist. We have a priori knowledge of logic because the human mind has the ability to grasp by rational intuition necessary truths about abstract reality.

The logical positivists were influenced significantly by Frege's criticisms of Kant and by Frege's logicist thesis about arithmetic. But there were certain aspects of Frege's position that they could not accept. For example, they could not tolerate the notion that any human knowledge could be synthetic a priori; thus they rejected the Kant-Frege line on geometry. Furthermore, the logical positivists found unacceptable any account of human knowledge, even our knowledge of logic, that involved some mystical form of quasi-Platonic rational intuition. Taking their cue from Wittgenstein's *Tractatus Logico-Philosophicus*, they regarded simple logical tautologies, not as expressing profound abstract truths into which our reason has some special insight, but as definitional truths resulting solely from the meanings of the terms and symbols used to express them. Logical and arithmetical statements are also true and known with certainty to be true precisely because they have no factual content: they are analytic statements; their truth is completely determined by the meaning of the words and symbols they contain. In this way, the positivists sought to explain the nature of arithmetical truth and the certainty of arithmetical knowledge without appealing to Platonic forms (such as numbers conceived of as abstract entities to which mathematical statements might be thought to correspond) or any form of mystical intuition (by which Platonists believe we acquire mathematical and logical knowledge).

It is unclear whether the logical positivists accepted the analyticity of arithmetic only because they also accepted Frege's logicist thesis or whether they would have continued to assert the analyticity of arithmetic even if the logicist thesis were to have been abandoned. As Hempel explains in "The Nature of Mathematical Truth," the logicist thesis ran into a number of severe problems.¹² One of these problems was the recognition that to reduce arithmetic to logic requires a liberal dose of set theory, and some of the axioms of set theory (such as the axiom of infinity and the axiom of choice) are neither tautologies nor reducible to tautologies by definitional substitution. Without the logicist thesis, the later logical positivists and empiricists were left with the definition of analyticity criticized by Quine: an analytic statement is one whose truth (or falsity) is completely determined by the meaning of the words and symbols it contains. This definition permits the truths of arithmetic to be analytic, and hence to be known a priori, even if arithmetic is not reducible to logic.

Concerning geometry, the logical positivists were struck by two things. First, the success of Einstein's general theory of relativity seemed to show that whether Euclidean geometry is true of space is an empirical matter, not something we can know a priori as Kant and Frege had thought. Second, despite the success of Einstein's theory, it still seemed that Euclidean geometry, along with its non-Euclidean rivals, could be studied as a legitimate branch of mathematics in which theorems are proved by deducing from them from axioms. For example, it is a nontrivial geometrical truth that we can demonstrate, independently of experience, that in Euclidean geometry (but not in alternative geometries) the sum of the interior angles of a triangle is exactly 180 degrees. The solution, they thought, was to distinguish *pure* geometry from *empirical* (or *interpreted*) geometry. Pure geometry, which explores the logical consequences of various axiom sets and the logical relations between axiom sets, is analytic and its results, like those of arithmetic, are known independently of experience. Empirical geometry, which involves hypotheses about the nature of physical space that can only be verified or falsified through observation and experiment, is synthetic. For example, once we have specified a physical interpretation of the term *straight line*—as, say, the path of a light ray in a vacuum—it is an empirical hypothesis that straight lines obey Euclid's fifth postulate concerning parallel lines.¹³

One of the principal motivations for adopting the analytic-synthetic distinction among philosophers of science in the first half of the twentieth century was to forge an account of mathematical truth and knowledge that is consistent with empiricism. All the necessary truths of arithmetic and pure geometry were judged to be analytic because their truth rests solely on linguistic conventions, on the meanings of words and symbols. The factual component of their meaning is nil. Thus, mathematical truths such as " $2 + 5 = 7$," and logical truths such as "everything that is both *F* and *G* is *F*," are in the same category as "all vixens are foxes" and "no sexa-

genarians are forty-five years old." All alike are analytic truths, true solely in virtue of the meanings of the words and symbols used to express them. And, because they are analytic, we can know them a priori, independently of experience.

Outside of pure mathematics and logic, synthetic statements abound. Unlike analytic statements, their truth depends on both a linguistic and a factual component. Obviously, the truth of a statement such as "cyanide is poisonous" depends on the meaning, in the English language, of the words *cyanide* and *poisonous*, and the rules governing the copula *is*. But no amount of linguistic analysis can tell us whether the statement is true. For that, we need experience and observation. The factual component is a function of the way the world is, and thus all synthetic statements are a posteriori.

According to the logical positivists, the meaning of each synthetic statement is given by its implications for experience and observation. This is usually referred to as the verifiability principle (or criterion) of meaning. The central idea is that an individual statement, such as "iron rusts in air" or "there are brick houses on Elm Street," has a cognitive or empirical meaning only if it logically implies a group of statements that are about our immediate experience or the outcome of observation. This is the doctrine that Quine calls *reductionism*: each individual synthetic statement has a meaning that is given by a unique set of experiential or observational consequences that follow from it and that, if true, would verify it conclusively. Synthetic statements that lack experiential or observational consequences are judged to be meaningless, devoid of cognitive and empirical significance. Into this disreputable category, the positivists threw many of the statements from theology and metaphysics. In later, slightly more liberal versions of reductionism, the meaning of a synthetic statement was identified with the observational or experiential statements that could either confirm or disconfirm it (even if they could not verify or refute it conclusively).

So the broad picture that emerges is that our knowledge of the truths of logic and pure mathematics is a priori because all such statements are analytic; they are true solely in virtue of the meanings of the words and symbols they contain. Ultimately, these meanings depend on the conventions of language. Statements that are not analytic are synthetic. All knowledge of synthetic statements is a posteriori: there is no synthetic a priori knowledge, as Kant and Frege had thought. Moreover, the meaning of each synthetic statement is given by the set of experiential sentences that it implies or the set of such sentences that could either confirm or disconfirm it.

QUINE'S REJECTION OF THE ANALYTIC-SYNTHETIC DISTINCTION

In the first four sections of "Two Dogmas" Quine criticizes the major attempts to define analyticity. After some brief remarks about Carnap's definition of analyticity in terms of state-descriptions, he concentrates on the Fregean definition, according to which a statement is analytic if it is a tautology or can be reduced to a tautology by means of definitions.¹⁴ The idea is simple enough: "No bachelor is married" is analytic because it can be reduced to the tautology "No unmarried man is married" by replacing *bachelor* with its definition, "unmarried man." In section 2 of his paper, Quine explains that, except when introducing brand new terms whose meaning we stipulate, definitions are acceptable only when they preserve the existing meanings of the terms in question: "unmarried man" defines *bachelor* (and thus can be substituted for it) only because the two expressions have the same meaning. So the search for an acceptable account of analyticity depends on our finding an acceptable account of synonymy (or sameness of meaning).

One approach to characterizing synonymy would regard two expressions as synonymous when they can be exchanged *salva veritate*, that is, without changing the truth or falsity of the sentences in which they occur. But as Quine argues in section 3 of his paper, such interchangeability, by itself, does not guarantee sameness of meaning. For "all and only bachelors are unmarried men" might be true for the same reason that "all creatures that have a heart are creatures that have kidneys" is true: the two expressions in each sentence have the same extension; they refer to exactly the same set of objects, but (at least in the case of hearts and kidneys) they do not mean the same thing. Something more is needed for the synonymy of X and Y, beyond the truth of the biconditional statement "X if and only if Y." We require not merely that the biconditional "X if and only if Y" is true but that it is necessarily true. But then, Quine argues, to assert "Necessarily, X if and only if Y" is to do nothing more or less than to assert that the sentence "X if and only if Y" is analytic, and thus that the terms X and Y are synonymous. So we have come full circle. In order to understand analyticity, we appeal to synonymy. To understand synonymy, we appeal to interchangeability *salva veritate*. But such interchangeability is not sufficient for sameness of meaning. So we must add the further condition that synonymous expressions be not only interchangeable *salva veritate* but necessarily so, and we are led back to analyticity. This attempt at characterizing synonymy has not led to an independent account of analyticity.¹⁵

In section 4 of his paper, Quine addresses another attempt to make clear the notion of analyticity. Perhaps the failure to give an independent characterization of synonymy (and the difficulty of deciding in many cases whether a given sentence is analytic) can be traced to the vagueness of

ordinary language. The remedy for vagueness is precision, and this is what artificial languages aim to provide. Quine is particularly concerned with the artificial languages invented by Carnap and others in which semantical rules for a given language generate all the analytic sentences in that language. But here again, Quine thinks that we encounter circularity, for he argues that when we ask what distinguishes these semantical rules from other semantical rules (such as those specifying all the truths of the language), the only answer that seems to be forthcoming is that *these* semantical rules are the ones that pick out all and only the analytic sentences. Once again, in our attempt to define analyticity we end up presupposing the very notion we wish to define. Thus (after some further objections to the semantical rules strategy that we shall ignore), Quine concludes that trying to understand analyticity by means of artificial languages is a "*feu follet par excellence*" (291); it is like chasing a will-o'-the-wisp that vanishes as soon as we approach it, only to reappear somewhere else.

Given the failure of several important attempts to give an adequate characterization of analyticity, Quine suspects that the whole analytic-synthetic distinction is a dogma, an unsupported and perhaps unsupportable article of faith. He reinforces this suspicion in section 5 of "Two Dogmas," where he criticizes the verifiability principle (and its cousin, the confirmability principle) as a criterion of meaning for synthetic sentences. Just as it proved difficult to justify the claim that some sentences (the analytic ones) are true solely in virtue of their meaning, so, too, does it prove difficult to characterize the meaning of each synthetic statement in terms of some special set of experiential or observational statements. In Quine's view, this kind of reductionism is based on the same kind of mistaken view about meaning as the doctrine of analyticity. Both dogmas incorrectly regard meanings as things that attach to words and to statements independently of the other statements that we accept. This leads us to assume, wrongly, that we can decide whether a given statement is analytic (in virtue of the meanings of the words and symbols it contains), or whether two synthetic statements mean the same thing (in virtue of their experiential or observational consequences), without considering the role that these words and sentences play in our entire web of belief.

QUINE'S HOLISM AND HIS REJECTION OF A PRIORI KNOWLEDGE

Given the complexity and abstractness of Quine's paper, it is easy to lose sight of the fact that Quine's primary interest is epistemological. Quine wants to deny that there is any such thing as a priori knowledge. He spends so much time attacking the doctrine of analyticity because his opponents (such as the logical positivists) had argued that we know some statements (such as logical and mathematical truths) to be true independently of

experience because they have the special semantical property of being analytic. If Quine can show that analyticity is a myth (an unsupported “dogma”), then one of the main supports of the epistemological thesis that we have a priori knowledge will have been destroyed.¹⁶

In the final section (6) of his paper, Quine sketches his own view of meaning and belief. Because Quine denies that any statement, even those in mathematics and logic, are a priori, he concludes that any of our beliefs could, in principle, be revised or abandoned in light of experience. Similarly, he contends here that any belief, even those traditionally regarded as synthetic, could be retained regardless of the outcome of observations and experiment. It is this aspect of Quine’s position that calls to mind Duhem’s holism and which has been the subject of considerable scrutiny by philosophers of science. Because Laudan discusses each of Quine’s arguments for holism in “Demystifying Underdetermination,” we shall defer consideration of those arguments until our discussion of Laudan’s article below.

3.3 | Gillies on Duhem and Quine

In “The Duhem Thesis and the Quine Thesis,” Donald Gillies explains the reasoning behind the versions of holism espoused by Duhem and Quine respectively. Gillies criticizes aspects of each and then concludes by advocating a modified version of a holist thesis that incorporates elements from both. We begin our discussion of Gillies’ article with a brief summary of three important differences he identifies between the theses of Duhem and Quine.

THREE CONTRASTS BETWEEN THE DUHEM THESIS AND THE QUINE THESIS

One major difference between Duhem and Quine is the scope of their holist claims. Duhem restricted his thesis regarding the ambiguity of falsification to theories in physics, excluding less abstract sciences such as physiology and botany. Quine, notoriously, extends holism to any sentence whatever—“the totality of our so-called knowledge or beliefs” (296)—thus including not only all the empirical sciences, from the most to the least abstract, but also logic and mathematics. Specifically, in “Two Dogmas” Quine claims that, (i) despite all the evidence we might accumulate, any statement whatever could be retained as true in the face of that evidence, and (ii) any sentence, even one from logic or mathematics, might be rejected as false.¹⁷

Closely related to the scope of their respective claims are their differing views concerning what Quine calls “the unit of empirical significance”

(296), the smallest linguistic grouping that logically implies observational statements. Just like the logical positivists he criticizes, Quine ties meaning (empirical significance, empirical content) to observational consequences and regards observational statements as unproblematic as far as their meaning is concerned. But unlike the logical positivists, who took the empirically significant unit to be the individual sentence of a scientific theory, for Quine the unit is nothing less than the whole of science. Anything smaller, in his view, lacks empirical significance because it fails to imply anything about observation. Duhem's view is far less sweeping. As we have seen, Duhem regards some individual sentences from low-level sciences as empirically significant in their own right. But in abstract, high-level sciences such as physics, he considers the smallest empirically significant unit to be a group of theories, hypotheses, and assumptions from within that science. The important contrast is that, even when espousing holism with respect to theories in physics, Duhem's brand of holism is far less global than Quine's: for Duhem, even the largest unit of empirical significance (in physics) is far smaller than the whole of science or even the whole of physics.

Another contrast between Duhem and Quine concerns ways of saving a theory (or a single theoretical statement) from falsification by modifying auxiliary hypotheses and assumptions. While both Duhem and Quine agree that deductive logic alone cannot compel a scientist to reject a theory as false, Duhem emphasizes that beyond a certain point, trying to retain an ailing theory in the face of predictive failures would be unreasonable and contrary to scientific good sense. For his part, Quine is content merely to stress the impotence of deductive logic to force such a decision. Although Quine mentions that pragmatic factors will play a role, those factors are not described in any detail.

GILLIES' CRITICISMS OF QUINE

Gillies' main criticism of Quine concerns Quine's sweeping claim that "the whole of science" (296) is the unit of empirical significance. As Gillies remarks, this claim is altogether implausible. When one examines cases in which a theory in physics (such as Newton's theory of gravitation) is combined with other physical theories and assumptions to generate observational predictions, theories from other sciences (such as biology, geology, and botany) play no role whatever in that derivation. On this issue at least, Duhem is closer to the truth when he restricts the scope of his holist thesis to a relatively small group of auxiliary hypotheses.

Gillies is concerned solely with the claims that Quine advanced in "Two Dogmas," but it is interesting to note that, in his later publications, Quine significantly tones down his thesis that the unit of empirical significance is the whole of science. Quine still insists that, in virtue of their use of logic and mathematics, the sciences are far more unified and in-

tegrated than is commonly realized, but he admits that "little is gained by saying that the unit is in principle the whole of science, however defensible this claim may be in a legalistic way."¹⁸

Another shortcoming of Quine's "Two Dogmas," according to Gillies, is Quine's almost total lack of interest in exploring ways in which the retention of a theory in the face of adverse evidence might be unreasonable even though it is not forced by deductive logic. (The arrow of *modus tollens* points at a whole group of theories, hypotheses, and assumptions without singling out the culprit.) In this respect, Duhem's insistence on the importance of good sense in science is a step in the right direction. (Although, as Gillies points out, Duhem himself seems to have been singularly lacking in such good sense, judging by the theories he rejected—relativity theory, the kinetic theory of gases, and the atomic theory.)

Finally, Gillies rejects Quine's contention that no scientific statement, not even experimental laws, can be tested (and thus falsified) without the aid of auxiliary hypotheses. Gillies thinks that within science, and even within physics, what he calls level 1 hypotheses are both confirmable and falsifiable. This claim will be assessed in the later section on Gillies' version of the Duhem-Quine thesis.

GILLIES' CRITICISMS OF DUHEM

Despite his criticisms of Quine, on one important matter Gillies thinks that Quine is right and Duhem wrong, namely, Quine's insistence that all of science, not just physics, falls within the scope of the holist thesis. There are really three aspects to this issue, which we will discuss in turn: the extension of the holist thesis to other sciences, mathematics, and logic.

OTHER SCIENCES

Gillies faults Duhem in two respects: for denying that there are any falsifiable laws in physics and, more importantly, for limiting his thesis exclusively to physics. Postponing the question of falsifiable laws to the following section on Gillies' version of the Duhem-Quine thesis, let us focus on the charge that Duhem incorrectly limits the scope of his holist thesis to physics.

It is true that the title of section 1 of Duhem's chapter proclaims that "the experimental testing of a theory does not have the same logical simplicity in physics as in physiology" (257), but Duhem's discussion reveals that the underlying logic is the same for both, at least when measuring instruments are involved. Ambiguity of falsification is avoided in physiological and chemical experiments involving measuring instruments only because physiologists and chemists (allegedly) accept many auxiliary hypotheses as established truths on the presumed infallible authority of physicists. In short, physiologists take for granted what physicists know to be based on theory. Thus, the difference lies in the psychology of testing, not

in its logic. This remains consistent with Duhem's view (articulated in a section not excerpted here) that in some (perhaps many) experiments outside of physics, no measuring instruments are involved, and we can rely on what Duhem calls "a recital of concrete and obvious facts," without the need for any theoretical interpretation (*Aim and Structure*, 147).

In short, the logic of Duhem's argument leads to the inclusion of other sciences within the scope of his thesis, since the testing of laws and theories in those sciences requires the use of instruments and special apparatus. Perhaps Duhem's failure to appreciate that the ambiguity of falsification is not limited to physics was a consequence of a tendency among physicists (of whom Duhem was one) to underestimate the degree of theoretical abstraction involved in other sciences. In addition, Duhem might have thought that all the auxiliary hypotheses involved in the use of instruments and special apparatus belong to physics rather than to the sciences in whose service they are employed.

MATHEMATICS

As far as philosophers of science are concerned, there are two main branches of mathematics: geometry and arithmetic. As Gillies points out, Duhem was convinced that Euclidean geometry had been conclusively established as true by our commonsense knowledge of the world. For example, Euclidean geometry (in which space is perfectly flat, having a constant, zero curvature) is the only geometry in which there exist similar figures of different sizes. In spaces of nonzero curvature, geometrical figures are similar if and only if they are equivalent (i.e., if and only if all their angles and sides have exactly the same magnitude). To Duhem, it seemed incontestable that simple, everyday observation reveals space to be Euclidean because of the "obvious" fact that there are similar figures of different sizes. Thus, Duhem summarily rejected the theory of general relativity—according to which the curvature of space varies—as a purely formal theory that could have no application to the real world. (Gillies conjectures that Duhem's rejection of relativity theory was colored by Duhem's fervent anti-German nationalism in the midst of the First World War.) A careful application of Duhem's own analysis of the logic of testing should have persuaded him that our everyday observations are consistent with the hypothesis that, while space has a small but nonzero curvature in the vicinity of the earth, the curvature might be greater elsewhere. Thus, Duhem was wrong to regard geometry (*applied* geometry) as immune from revision. Gillies says nothing about Duhem's views concerning arithmetic, but presumably Duhem was equally conservative about that as well. Even if one is not a logicist (one who believes that arithmetic can be reduced to logic and set theory), arithmetic seems to be far more intimately related to logic than geometry is. Thus, it is plausible to think that arithmetic is revisable only if logic is, and it is to the status of logic that we now turn.

LOGIC

Duhem was unaware of the modern system of logic developed by Frege, Bertrand Russell, and Alfred North Whitehead, believing as he did that logic began and ended with Aristotle. This ignorance, by itself, is not terribly significant because the new logic is not inconsistent with Aristotle's theory of the categorical syllogism (at least when the modern view that universal statements lack existential import is adopted). The new logic simply includes Aristotle's theory as a small set of special cases in a much wider system. What Duhem might be faulted for is his uncritical assumption that the laws of logic are immune from revision or rejection. Of particular interest here is the law of excluded middle.

In modern terms, the law of excluded middle says that, for any statement p , the statement p or *not- p* is a logically necessary truth: it remains true under all possible assignments of truth value to p . The law of excluded middle is closely related to another logical law, the law of bivalence. Understanding the subtle difference between these two laws will help us to appreciate what is at stake when people urge us to abandon the law of excluded middle. The law of bivalence says that every statement has a truth value and that there are just two truth values: truth and falsity. In other words, the law of bivalence asserts that, for any statement p , either p is true or p is false. Most of the attacks on the law of excluded middle have come from those who reject bivalence. In systems of many-valued logic (such as those developed by the Polish logician, Jan Lukasiewicz), there are more than two truth values. Suppose there are three truth values, 1, 2, and 3. If we identify 1 with truth and 3 with falsity, then we have to give up the law of excluded middle because it then becomes possible that statement p is neither true nor false but has truth value 2 (which we might call *indeterminate* or *intermediate*.) Alternatively, if we wish to preserve the law of excluded middle by identifying falsity, not with 3, but with either 2 or 3, then we have give up one of the other laws of classical logic, such as the law of double negation. To see why this is so, consider the requirement that negation be a truth-functional operator: it must always lead from one particular truth value to another particular truth value. Suppose that negation always takes us from 1 to 3. Then, 2 cannot also lead to 3 but must lead to 1. Similarly, given these assignments, negation must take us from 3 to 2. Thus, double negation will take us from 1 to 3 to 2 rather than back to 1. So the fundamental law of classical logic, that for any statement p , p is logically equivalent to *not-not- p* , must be given up.

As we have seen, the real issue in rejecting the law of excluded middle is whether we should abandon the classical law of bivalence, that every statement is either true or false. For we could, if we wished, reject bivalence, and retain the law of excluded middle, but then we would have to abandon at least one of the other laws of classical logic such as double negation. As Quine has remarked, in his *Philosophy of Logic* (1970), "it

is hard to face up to the rejection of anything so basic."¹⁹ Gillies notes that two main reasons have been given for adopting some form of deviant logic: intuitionism and quantum mechanics. Intuitionism is a movement in the philosophy of mathematics founded by the Dutch mathematician Luitzen Egbertus Jan Brouwer (1881–1966). Brouwer was suspicious of mathematical arguments that profess to prove claims about infinite sets (such as the set of all numbers) without actually giving us a method for constructing an instance of that claim for any given member of the set (such as an arbitrarily chosen number). The arguments that Brouwer rejected are *reductio* arguments relying crucially on the law of double negation. As far as Brouwer was concerned, *reductio* arguments were not proofs because proofs have to be constructive. (For obvious reasons, intuitionism is often called constructivism.) Thus, Brouwer rejected the law of double negation. Brouwer also rejected the law of excluded middle because he identified truth with (constructive) proof. Not every mathematical proposition is such that either we can prove that proposition or we can prove its negation. Consequently, because some statements (as, for example, Goldbach's conjecture that every even number greater than two is the sum of two primes) are undecided, the principle that, for all p , p or *not- p* is a necessary truth is not a law of intuitionist logic. In this way, intuitionists deny that every proposition is either true or false.

Intuitionist logic lacks the simplicity and elegance of classical logic. (Among other things, none of the connectives in intuitionist logic can be defined in terms of the others.) Its predominant feature is that it reduces the number of propositions that can be deduced from others. The question why we should accept this limitation on what constitutes acceptable reasoning naturally arises. It should be clear from our brief sketch that the rationale for intuitionism has nothing to do with observation, prediction, or anything empirical, the considerations that normally apply to purported instances of the Duhem-Quine thesis. Rather, the case for intuitionism rests on the relationship between logic and mathematics and on a highly controversial—and to many, implausible and idealist—thesis about the status of mathematical objects (such as numbers and geometrical figures).

Logic is widely regarded as the study of the general principles of reasoning that apply to all arguments, regardless of their subject matter. Intuitionists disagree. They claim that logic is, fundamentally, the set of principles that govern mathematical reasoning and that we discover what these principles are by studying mathematics. Thus the intuitionists' repudiation of classical logic depends on their distinctive views about mathematics.

According to the intuitionists, mathematical objects do not exist independently of the human mind. They are not (so to speak) out there in the real world, waiting to be discovered, nor do objective truths about them exist independently of human thought. Rather, numbers and the like are said to be mental constructs, the result of mental operations per-

formed by human beings. Mathematical objects exist if and only if they can be constructed; a mathematical statement is true if and only if it has a (constructive) proof. (Because it collapses the distinction between truth and provability and denies that it even makes sense to talk about a mathematical statement that is true but has no proof, intuitionism is a variety of verificationism.)

Quantum logic was first proposed by the mathematicians Garrett Birkhoff and John von Neumann in 1936 as a solution to the paradoxes of quantum mechanics. Quine alludes to it in "Two Dogmas" as a possible instance of the revisability of logic, and, more recently, Hilary Putnam has championed it as showing that logic is empirical. The essential feature of quantum logic is that it gives up the classical distribution laws for conjunction and disjunction. In standard classical logic, (1) logically implies (2).

$$1 \quad X \& (Y_1 \vee Y_2).$$

$$2 \quad (X \& Y_1) \vee (X \& Y_2).$$

In quantum logic, the inference from (1) to (2) is invalid. Take, for example, the classic two-slit experiment in which electrons are fired at a plate with two slits in it and then form an interference pattern on a screen. Consider a single electron passing through the apparatus, and let X = "The electron is in region R on the screen," Y_1 = "The electron went through slit number 1," and Y_2 = "The electron went through slit number 2." Part of the oddity or "paradoxical" character of quantum systems and of quantum mechanics that correctly describes them is that while (1) is true, (2) is false. The pattern we obtain on the screen is typical of the interference pattern we would obtain if, instead of electrons, we had used light waves (or even water waves). If the electron had gone either through slit 1 (but not slit 2) or through slit 2 (but not slit 1), no interference pattern would have been formed.²⁰ To get the interference pattern, each single electron must act like a coherent wave front that passes through both slits and interferes with itself! To its supporters, the attraction of quantum logic is that it does not permit inferences to any conclusion that quantum mechanics and experiment reveal to be either false or unverifiable; in this way the paradoxes of quantum behavior are avoided. To its detractors, quantum logic does not solve any paradoxes—it simply shifts the mystery from physics to logic. Very few physicists or philosophers have argued that we should change the way that we reason about the world—a world that includes not just electrons and neutrinos but tables, chairs, and kangaroos—by giving up the distribution laws for disjunction and conjunction.²¹

It is interesting to note that in his *Philosophy of Logic*, Quine argues that whenever someone tries to deny a law of classical logic such as excluded middle, the person merely "changes the subject" by changing the

meanings of the logical connectives (83). This is a far cry from the Quine of "Two Dogmas" for whom any rule of logic could, in principle, be rejected as false.²² The problem to which Quine (in 1970) is drawing attention is that neither intuitionist logic nor quantum logic is truth functional: in neither case can we give truth tables for all the connectives using just the two truth values, truth and falsity. Thus, we cannot appeal to truth tables to decide whether the symbols $\&$, \vee , \sim , and \supset denote the same concepts of conjunction, disjunction, negation, and implication in deviant logic as they do in classical logic. The deviant logics are presented axiomatically and lack some of the axioms of classical logic. Thus, someone like Quine (in 1970) could insist that when the axioms are changed so, too, are the connectives that the axioms implicitly define.²³

In summary we can say that a good case can be made for extending Duhem's thesis to sciences other than physics and for including geometry within its scope. But it is far less clear that we should follow Quine's "Two Dogmas" and Gillies by also including arithmetic and logic. Certainly no one, not even Putnam, has suggested that every law of classical logic could, in principle, be rejected. In fact, Putnam has argued that at least one logical law is a genuine a priori necessity, namely the law of noncontradiction.²⁴

GILLIES' VERSION OF THE DUHEM-QUINE THESIS

Gillies concludes his article by offering a revised version of the Duhem-Quine thesis that he considers both true and important. In order to clarify the relationship between Gillies' thesis and the views of Duhem and Quine, we have followed Gillies in dividing the thesis into several distinct parts. (The only difference between our division and Gillies' is that we have split Gillies' second clause into the two distinct claims, B and C.)

- A "The holist thesis applies to any high-level (level 2) theoretical hypotheses, whether of physics or of other sciences, or even of mathematics and logic." (317) (Contrary to Duhem, in light of Quine.)
- B "The group of hypotheses under test in any given situation is in practice limited, and does not extend to the whole of human knowledge." (317) (Contrary to Quine, in light of Duhem.)
- C "Quine's claim that 'Any statement can be held to be true come what may' . . . is true from a purely logical point of view; but scientific good sense concludes in many situations that it would be perfectly unreasonable to hold on to particular statements." (317) (Addition to Quine, in light of Duhem.)

Since we have already discussed *B* and *C*, we shall conclude this section on Gillies by taking a brief look at *A*.

An important feature of *A* is Gillies' restriction of the holist thesis to what he calls level 2 hypotheses. Earlier in his article, Gillies distinguishes between high-level (level 2) theoretical hypotheses and low-level (level 1) generalizations such as experimental laws. Gillies describes level 1 generalizations as those that can be falsified and confirmed directly by observation and experiment. Thus, Gillies thinks that Duhem's minimal thesis about falsification holds only when restricted to level 2 hypotheses (and higher). But clearly, Gillies runs the risk of trivializing the holist thesis if level 2 hypotheses are simply defined as those that satisfy the thesis and level 1 hypotheses are defined as those that do not. Hence, we need some independent characterization of the two types of hypothesis.

In his book *Philosophy of Science in the Twentieth Century*, from which Gillies' article is taken, he gives as examples of level 1 hypotheses Snell's law of refraction applied to glass, Kepler's first law, Planck's radiation law for black body radiation, and Einstein's equation for the photoelectric effect. Gillies' examples of level 2 hypotheses include Newton's first law of motion and Einstein's quantum theory of electromagnetic radiation. What Gillies seems to have in mind as level 1 hypotheses are what are often called *descriptive laws*, that is, generalizations using only variables such as position, velocity, temperature, and wavelength, whose values can be measured with instruments. As the name *descriptive laws* suggests, level 1 hypotheses are confined to describing the phenomena. Unlike level 2 hypotheses, level 1 hypotheses do not attempt to explain the phenomena they describe, nor do they deal with causes. But even so, this hardly establishes that level 1 hypotheses can be confirmed or falsified without using any theoretical assumptions whatever. For as long as measuring instruments are involved, or there is a need to verify that disturbing influences are absent, theories from other branches of physics will have to be used. As we indicated earlier, this is Duhem's central argument for his thesis of the ambiguity of falsification. Gillies appears to have simply assumed that because a law is descriptive (as opposed to being causal or explanatory) it can be tested without using any theories whatever. On this issue, at least, Duhem seems to have been closer to the truth.

3.4 | Laudan's Repudiation of Underdetermination

In "Demystifying Underdetermination" Larry Laudan takes aim at a variety of underdetermination theses that have become fashionable in the philosophy of science and epistemology since the publication of Quine's "Two Dogmas." As Laudan notes in his introduction, it is often unclear what specific claim about underdetermination a particular author is espousing

and, partly as a consequence of this unclarity, relativistic consequences have been drawn from the presumed fact of underdetermination, vaguely construed, that do not follow from some of its more specific versions. Of particular concern to Laudan is the tactic of bait-and-switch: one is lured into accepting a fairly plausible version of underdetermination (the consequences of which are innocuous) only to discover that conclusions are then drawn that follow only from another, far less plausible version of underdetermination (the consequences of which are devastating).

Laudan's goal, then, is threefold: to clarify the different sorts of claims about underdetermination, to assess their plausibility, and to see what follows from them. To this end Laudan divides claims about underdetermination into two classes: the deductive and the ampliative. Deductive claims limit themselves to what can be established about the status of theories, given some evidence, using only deductive logic. Ampliative claims permit the use of nondeductive inferences as well. Nondeductive inferences are involved in judging how well the available evidence confirms a theory, whether one theory is more strongly confirmed by the evidence than a rival theory, and whether one theory explains the evidence better than its competitors. The classic example of a deductive thesis concerning underdetermination is the one associated with David Hume. All of the more recent versions of underdetermination, beginning with Quine and extending through Goodman, Kuhn, Hesse, and Bloor, are ampliative.²⁵

One important preliminary matter needs to be addressed before proceeding to the heart of Laudan's article. This concerns the concessions that Laudan is prepared to make to the philosophers whom he criticizes. To avoid needless debate, Laudan is willing to adopt an extremely broad notion of what constitutes a theory. He says that he will regard as a theory any set of universal statements that purport to describe the world. Presumably, he means that a theory should be empirical (it should make a definite claim about the world that could, logically, be false) and that it should contain at least one universal generalization (rather than, as his wording suggests, requiring that a theory contain nothing but universal statements). In addition, Laudan is willing to grant that single theories by themselves do not make any directly testable assertions. This, we will assume, means that Laudan accepts Duhem's holist theses (D1) and (D2): single theories yield testable predictions only in conjunction with auxiliary hypotheses and assumptions.

HUMEAN UNDERDETERMINATION (HUD)

Laudan defines deductive or Humean underdetermination as follows:

HUD For any finite body of evidence, there are indefinitely many mutually contrary theories, each of which logically entails that evidence. (323)

Laudan comments that the arguments for HUD are familiar and trivial, and that HUD, while true, merely shows that the fallacy of affirming the consequent is indeed a fallacy: the mere fact that a particular theory (T_1) entails some evidence (E) does not logically imply that T_1 is true, since indefinitely many rival theories, T_2 , T_3 , and so on, have this same property. The weakness of HUD, in Laudan's view, is that it restricts itself solely to deductive logic, completely ignoring the possibility that some kind of ampliative inference (involving, say, inductive confirmation) might single out T_1 as more probable than its rivals. Moreover, even if one were to ignore this possibility and appeal to HUD to license some kind of relativist conclusion, at best what would follow is that all theories that entail E are equally worthy of belief, not that it would be rational to accept any theory whatever, regardless of its relation to E .

What are the trivial and familiar arguments for HUD to which Laudan refers? One common argument appeals to the logical possibility of generating indefinitely many theories, each of which logically entails E , simply by conjoining E with each member of a set of mutually contrary statements. Thus, we could have:

- T_1 ($A \ \& \ E$)
- T_2 ($B \ \& \ E$)
- T_3 ($C \ \& \ E$), and so on,

where, for example, A says that all electrons have a mass of exactly one gram, B says that all electrons have a mass of exactly two grams, and so on. Another trivial strategy involves using the truth-functional conditional connective \supset in the following way:

- H_1 ($A \supset E$) & A
- H_2 ($B \supset E$) & B
- H_3 ($C \supset E$) & C , and so on.

Of course, no scientist would give these so-called theories the time of day, either because they are already known to be false or because they are utterly lacking in explanatory power and predictive capacity. But that is not the point. HUD is concerned solely with what is logically possible. Given a sufficiently liberal notion of what qualifies as a theory, these examples show that HUD is true—true but profoundly uninteresting.

There is one small problem with the wording of HUD, given the context of Laudan's article. Laudan is prepared to concede, for the sake of argument, that Duhem was right and that no single theory, by itself, makes any directly testable assertion. But HUD says that for any evidence, E , indefinitely many theories logically entail E , and in many cases, E will be a directly testable assertion. So HUD is inconsistent with Laudan's concession. The problem is easily remedied if we follow the suggestion in

the appendix to Laudan's article (346–47) and rephrase HUD slightly as HUD*.

HUD* For any finite body of evidence, there are indefinitely many mutually contrary theories such that each theory, in conjunction with appropriate auxiliary hypotheses and assumptions, logically entails that evidence

QUINEAN UNDERDETERMINATION (QUD)

Since Quine is one of the most influential proponents of underdetermination in the twentieth century, he is Laudan's main target and receives the lion's share of attention. Laudan begins by attributing to Quine two doctrines: the nonuniqueness thesis and the egalitarian thesis. (For convenience, we shall refer to these using the acronyms NUT and EGAL.)

NUT "For any theory, *T*, and any given body of evidence supporting *T*, there is at least one rival (i.e., contrary) to *T* that is as well supported (by that evidence) as *T*" (324)

EGAL "Every theory is as well supported by the evidence as any of its rivals" (324)

As Laudan notes, both of these doctrines are epistemic: both involve the notion of empirical support, which is a type of evidential warrant or justification—the sort of thing that, when there is enough of it, makes belief rational. In Laudan's terminology, both NUT and EGAL are varieties of ampliative underdetermination. Clearly, EGAL is the stronger claim: EGAL entails NUT, but not vice versa. Laudan claims that Quine is committed to both doctrines—explicitly to NUT and implicitly to EGAL, the sweeping doctrine that strong relativists take Quine to have made plausible in "Two Dogmas." Laudan thinks that the relativists are half right: right in attributing EGAL to Quine, wrong in thinking that Quine has done anything to render EGAL plausible.

Laudan's account of Quine's relation to EGAL is rather complicated. Laudan argues that the doctrines that Quine holds explicitly, in particular Quine's claim that

(0) one may hold onto any theory whatever in the face of any evidence whatever (325),

implicitly commit Quine to EGAL. In Laudan's words, Quine's explicit endorsements of (0) "presuppose the egalitarian thesis, and make no sense without it" (324). Thus, Laudan is claiming not only that (0), when cor-

rectly interpreted, implies EGAL, but also, and more importantly, that unless Quine can show that EGAL is plausible, his case for (0) crumbles. It is important to realize that Laudan's second claim does not follow logically from his first claim but must be established on independent grounds. In general, one can be justified in holding a belief without first having to be justified in believing everything that is entailed by the belief. In the typical case of belief formation, one first believes some proposition on what one takes to be good grounds, and then one acquires a justified belief in the further propositions that follow from it. To insist that, in order to be justified in believing one proposition, one must first be justified in believing all the logical consequences of that proposition, is to place an unmeetable and hence unreasonable demand on what it takes for a belief to be justified. So Laudan must show that Quine's case for (0) vitally depends on accepting EGAL; Laudan cannot simply appeal to the fact that (0) has EGAL as one of its consequences.²⁶

Laudan's account of why (0) depends on EGAL runs as follows: First, he notes that, in order for (0) to be at all philosophically interesting, it has to be read normatively, not descriptively. The issue is not whether people can, by whatever means, retain their belief in any theory come what may, but whether they can do so rationally. So, Laudan interprets (0) as (1):

- (1) It is rational to hold onto any theory whatever in the face of any evidence whatever. (325)

Has Quine established (1)? Laudan thinks that in order to do so, Quine would have to examine every single ampliative rule governing rational theory choice and show that none of them has the power to favor one theory over another. In fact, the only rule that Quine considers is the Popperian admonition to "reject theories that have (known) falsifying instances" (326).

Before discussing Laudan's criticism of Quine's conclusion that the Popperian falsification rule radically underdetermines theory choice, we pause to consider whether Quine is as negligent as Laudan charges in ignoring all the other rules governing the choice among rival theories. Someone interested in defending Quine might point out that Quine's concern is not whether we should accept a given theory, but whether evidence could make it irrational for us to continue to accept that theory, however irrationally we might have arrived at that theory in the first place. Thus, one might argue that ampliative rules concerning when it is rational to accept a theory as well confirmed by evidence are not relevant to Quine's inquiry and that he is justified in restricting his attention to the Popperian rule for theory rejection. But this attempt to defend Quine fails for at least two reasons. First, rules for theory acceptance clearly are relevant to theory rejection, for if one is rationally justified in accepting

theory T_2 , then one is justified in rejecting as false all the contraries of T_2 . The fact that T_1 was there first, so to speak, should count for little if the rule says that, in light of evidence E , T_2 is the theory one should accept and T_1 is inconsistent with T_2 . Second, there might be rules concerning theory rejection other than the Popperian one. For example, a rule might say that if several rival theories entail exactly the same evidence, then rationality requires one to reject as false the theory having the lowest predictive content, the least explanatory power, and the lowest degree of simplicity (if there is any theory that scores lowest in all three areas). The point is not to advocate such a rule as correct but merely to remind ourselves that the Popperian falsification rule—essentially, *modus tollens*—is not the only plausible rejection rule once we include ampliative inferences that go beyond deductive logic. Thus Laudan seems to be correct: there is a significant gap in Quine's argument when, in defense of (1), he confines himself solely to arguing for the inconclusive nature of Popper's rejection rule.

Even though Quine confines his attention solely to Popper's rejection rule, some of his attempts to defend (1) in the face of that rule are, as Laudan says, "pretty trifling stuff" (327). The possibility of changing the meanings of key words, of abandoning the laws of logic, and of pleading hallucination do not come close to showing that any theory can be retained rationally in the face of any evidence. Insofar as Quine has a case for (1), it rests entirely on his doctrine of underdetermination (QUD).

QUD Any theory can be reconciled with any recalcitrant evidence by making suitable adjustments in our other assumptions about nature. (328)

Now, as Laudan points out, in order for QUD to be in any way relevant to (1), it must concern ways in which we can rationally reconcile any theory with any evidence. If QUD is talking merely about what is logically possible, however irrational, then Quinean underdetermination is an utterly trivial consequence of holism.²⁷ So we must add the crucial adverb *rationally* to modify the verb *reconcile*. Moreover, the notion of reconciliation is multiply ambiguous. Laudan distinguishes four senses of what it might mean to reconcile a theory with recalcitrant evidence. The theory (or, more precisely, the group consisting of the theory together with auxiliary hypotheses and assumptions) may:

- be logically compatible with the evidence;
- logically entail the evidence;
- explain the evidence;
- be empirically supported by the evidence. (329)

Which one of these does Quine have in mind? We can gain further insight into Laudan's criticism of QUD by considering the formal scheme we introduced earlier in this commentary, in the section on Duhem (pp. 355-65). Let the theory whose rational retention is at issue be T . We grant, along with Duhem, Quine, and Laudan, that T by itself does not yield testable predictions. For this we need auxiliary hypotheses and assumptions. Let these auxiliaries be A_1, A_2, \dots, A_n , and let the testable prediction made by the entire group ($T \& A_1 \& A_2 \& \dots \& A_n$) be O_1 . When we perform the experiment or make the observation, what we actually find is not O_1 but E , where E entails that O_1 is false. Schematically we have:

$$(T \& A_1 \& A_2 \& \dots \& A_n) \rightarrow O_1,$$

E , and $(E \rightarrow \sim O_1)$.

We can now express relatively precise versions of Quine's underdetermination thesis, corresponding to the four different meanings of *reconcile* on Laudan's list. Here are the first two:

QUD1 Any theory, T , can be rationally reconciled with any recalcitrant evidence, E , by deleting some of the original auxiliaries, and perhaps adding a new auxiliary, B , such that the new group ($T \& B \& A_2 \& \dots \& A_n$) does not entail anything that is inconsistent with E .

QUD2 Any theory, T , can be rationally reconciled with any recalcitrant evidence, E , by deleting some of the original auxiliaries, and adding a new auxiliary, B , such that the new group ($T \& B \& A_2 \& \dots \& A_n$) entails E .

Laudan makes four important points about QUD1 and QUD2.

First, neither consistency with E nor entailment of E guarantees that the new group will explain E or that the new group will be empirically supported by E . (Whether theories are automatically confirmed by their deductive consequences is addressed in chapter 4; explanation is discussed in chapter 6. We will here simply assume that Laudan is right about this: deductive entailment is sufficient neither for explanation nor for confirmation.) Second, neither Quine nor anyone else has proven that QUD2 is true. In other words, no one has shown that there will always be rationally acceptable auxiliaries such that, when added to any theory T , the entire group will entail any evidential statement. Third, deleting some of the original auxiliaries and perhaps adding the new auxiliary, B , does not guarantee that the new group will have the same degree of explanatory power or the same degree of empirical support as the original group. Indeed, the deletion of some of the old auxiliaries will not only prevent the derivation of O_1 , but is also likely to diminish the predictive and explan-

atory power of the new group as compared with the old. Fourth, the shift from the old group to the new group is rational only if the new group has a significant degree of empirical support. (Laudan also considers the role of other theoretical virtues, such as simplicity and explanatory power, in making theoretical changes rational.)

So Laudan returns the Scottish verdict, not proven, in response to Quine's underdetermination thesis (with more than a hint of skepticism regarding its truth). Quine could render his thesis plausible only by giving a convincing argument for the egalitarian thesis, EGAL. For only if every theory (or, more relevantly, every group consisting of a theory together with auxiliary hypotheses and assumptions) enjoyed the same degree of empirical support, could any theory whatever be rationally retained in the face of any evidence whatever. But Quine does not argue for EGAL. He simply assumes that EGAL is true without giving any argument that renders it plausible.

In the remainder of his section on Quine, Laudan considers whether Quine has made a better case for underdetermination in his later book, *Word and Object*. Laudan's judgment, again, is negative. He concludes that *Word and Object* merely provides poor arguments for the weaker thesis of nonuniqueness (NUT). Since NUT does not entail EGAL, and it is EGAL (not NUT) that is required for any strong and interesting version of relativism, devotees of underdetermination need to look elsewhere for support. This search for more convincing support for EGAL leads Laudan to consider the views of Goodman, Kuhn, Hesse, and Bloor in the remainder of his article.

GOODMAN'S GRUESOME NEW RIDDLE OF INDUCTION

The predicate *grue* first made its appearance in the third chapter of Nelson Goodman's *Fact, Fiction, and Forecast* (1955). Goodman defined this decidedly weird predicate as follows: an object is *grue* if and only if either the object is examined before time *T* and is green, or the object is not examined before time *T* and is blue.²⁸ Goodman's purpose in introducing this gerrymandered predicate was to make a telling point about induction and inductive confirmation. Conventional wisdom has it that a generalization such as "All emeralds are green" is confirmed by its instances. We inspect one emerald and find that it is green. We observe another and find that it, too, is green. Soon, we infer from this evidence, that all *observed* emeralds are green, that it is very probable that *all* emeralds are green, whether we have observed them or not; and we predict further that the emeralds we will observe in the future will also be green. But if all the green emeralds that we have inspected were examined before time *T* (the year 2000, say), then it is equally true to say that all the emeralds we have observed have been *grue*. And if generalizations are confirmed by their instances, then the generalization that all emeralds are *grue* is

also confirmed by our evidence. But if all emeralds are *grue*, then we predict that any emerald examined after time *T* will not be green but blue. The straight hypothesis (that all emeralds are green) and the bent hypothesis (that all emeralds are *grue*) are genuine rivals that make differing predictions about the future. Nonetheless, if generalizations are confirmed by their instances, both hypotheses are confirmed by our evidence. Goodman calls this the new riddle of induction. The riddle is not to explain why we are rationally justified in accepting any generalization as true (or probable) when we have observed several instances of it and no counterinstances: that is the *old* riddle of induction, Hume's problem of justifying induction. Rather, the *new* riddle of induction is to explain why some generalizations (the straight hypotheses) are confirmed by their instances and others (the bent hypotheses) are not.

There have been many attempts to solve Goodman's new riddle of induction, including Goodman's own theory of entrenchment. Very roughly, Goodman argues that hypotheses are confirmed by their instances only when they use predicates that are well entrenched, and predicates become entrenched when they have been used to make successful inductive inferences in the past.²⁹ Other authors have argued that there is a relevant epistemological asymmetry between the predicates *grue* and *green* because, unlike *green*, *grue* is positional: it makes reference to a particular time. Hence, while we can judge the colors of objects by simple inspection, we cannot decide whether an object is *grue* without knowing the date of our observation. And this, some claim, is what makes bent hypotheses illegitimate as candidates for induction, not the entrenchment of predicates.³⁰ Other philosophers deny that generalizations are in fact always confirmed by their instances; Popperians, for example, argue that hypotheses are confirmed (or "corroborated") only when they survive severe tests, that is, tests that are designed to refute them.³¹ Finally, Bayesians insist that the solution to Goodman's problem lies in paying attention to the *degree* to which hypotheses are confirmed by their instances.³² Degree of confirmation depends, in part, on the prior probability of the hypothesis relative to our background knowledge. Our background knowledge gives a very low prior probability to bent hypotheses as compared with their straight rivals. Hence, a bent hypothesis receives very little confirmation from the evidence, indeed none at all if the prior probability of the bent hypothesis is zero. (See chapter 5 for a discussion of the Bayesian approach to confirmation theory.)

Laudan neatly sidesteps the debate over Goodman's new riddle of induction and its proposed solutions. Laudan is concerned solely with the egalitarian thesis that, given some evidence, every theory that entails the evidence is as well supported as its rivals. He makes two simple points. First, not even Goodman himself accepts that bent hypotheses receive the same inductive support as straight hypotheses from the observation of their instances. The whole point of the new riddle is to find out what principles

of ampliative inference should replace the simple, but flawed, principle of enumerative induction (that is, from all observed *As* are *B*, infer that all *As* are *B*). Second, even if the bent and straight hypotheses were equally well supported by the evidence, the egalitarian thesis would not be vindicated, since bent hypotheses are only a subset of all the possible alternatives to any given hypothesis. As Laudan reminds us, nothing in Goodman's new riddle of induction suggests that, when we have observed a number of green emeralds, the hypothesis that all emeralds are red is just as well confirmed as the hypothesis that all emeralds are green.

KUHN'S THESIS OF LOCAL UNDERDETERMINATION

Laudan's criticism of Kuhn's version of the underdetermination thesis is closely related to Laudan's attack on Kuhn in his "Dissecting the Holist Picture of Scientific Change" in chapter 2, and should be read in conjunction with that piece. Here, Laudan argues that Kuhn is implicitly committed to a restricted version of the egalitarian thesis, namely, that any theory can be shown to be as well supported by any evidence as any of its known rivals. This is the thesis designated (2*) earlier in Laudan's article. As Laudan notes, (2*) is weaker than the unrestricted egalitarian thesis we have called EGAL, since, unlike EGAL, (2*) refers only to the known rivals of a given theory, not to all possible rivals. In Laudan's terminology, Kuhn espouses a thesis of *local*, not *global*, underdetermination. Even when we restrict our attention to the relatively few rival theories competing for the allegiance of scientists at a given time, Kuhn's view is that shared methods and rules (what, in chapter 2, we called *cognitive* or *constitutive values*) are always insufficient to determine the choice. Only because he holds this view—the restricted egalitarian thesis (2*)—can Kuhn conclude that in all cases of theory choice the final decision must depend on subjective factors (what, in chapter 2, were called *contextual values*.)

Kuhn's argument for local underdetermination rests on his claim that the shared standards (or *values*, as Kuhn calls them) of accuracy, consistency, scope, simplicity, and fruitfulness are too vague to lead proponents of rival theories (or paradigms) to the same verdict. (See the section "Kuhn's Second Thoughts about the Rationality of Scientific Revolutions" in the commentary on chapter 2.) In response, Laudan concedes that the standards Kuhn has mentioned are vague and that, as a consequence, they are unlikely to point to a unique resolution that all scientists will accept. But, Laudan objects, Kuhn's list of shared standards does not include many of the rules that scientists commonly employ in choosing among rival theories. These rules include the admonition to prefer theories that are internally consistent³³ and the directive to prefer theories that make predictions that are surprising given our background assumptions. Unlike the items on Kuhn's list, these rules are quite precise. Certainly there is no

reason to think that they can never suffice to pick out one theory as better than its rivals. Hence, Kuhn's thesis of local underdetermination is unproven.

Laudan admits that some theoretical choices are underdetermined relative to some restricted body of evidence and some subset of shared methodological rules. What he strenuously denies is Kuhn's thesis that all theoretical choices are necessarily underdetermined by any body of evidence and the set of all such rules. To show the falsity of Kuhn's thesis of local underdetermination, Laudan considers in some detail one important historical case: the disagreement in the early eighteenth century between the Newtonians and the Cartesians about the shape of the earth.

In the seventeenth century, Descartes argued that the effects of gravity on a body near the earth's surface are due entirely to a swirling vortex of transparent matter that presses down on the earth. The axis of rotation of this vortex coincides with the earth's north-south axis. The force exerted by this vortex will be greatest at the equator and least at the poles. Hence, according to Descartes and his followers, the earth is compressed around the equator. So the Cartesians predicted that the earth's radius at the equator should be slightly less than the earth's radius at either pole. The Newtonians made the opposite prediction. On Newton's theory, the effects of gravity are due to mutual attraction between a body and the earth, and this force varies inversely as the square of their distance apart. As the earth rotates once every twenty-four hours about its north-south axis, centrifugal force will counteract this gravitational attraction. The opposing centrifugal force will be greatest at the equator and least at the poles. Hence, according to Newton and his followers, the earth should bulge at the equator; the earth's radius at the equator should be slightly greater than the earth's radius at either pole. The Académie des Sciences in Paris sent surveying expeditions north and south, to Lapland and Peru, to measure the length of a terrestrial degree at both locations. The results were unequivocally in favor of Newton and against Descartes: a degree at the earth's surface was longer in Peru than in Lapland. As Newton had predicted, the earth is fatter at the equator.³⁴ Both the French and the English agreed on this verdict. Because of the findings of the two expeditions, Newton's theory was accepted and Descartes' theory rejected.

Laudan does not claim that the French expeditions performed a crucial experiment in the sense that they established conclusively that Newton's theory is true and Descartes' false. His point is more modest but no less telling against Kuhn: here is an actual historical case, involving proponents of rival paradigms, the Newtonian paradigm and the Cartesian paradigm, in which a single, shared, methodological rule led to a decision that both sides accepted as correct. The rule in question is, roughly, that when two rival theories make conflicting predictions that can be tested in a manner that does not presuppose the truth of either theory, then one should accept the theory that makes the correct prediction and reject its

rival. One example cannot prove a universal generalization, but one counterexample can refute it. The case of the rival theories about the shape of the earth is a counterexample to Kuhn's thesis of local underdetermination. Hence, Kuhn's thesis is false.

THE STRONG PROGRAMME IN THE SOCIOLOGY OF SCIENCE

In his paper's last section, "Underdetermination and the 'Sociologizing of Epistemology,'" Laudan takes aim at Mary Hesse and David Bloor, who have argued that underdetermination implies that scientists' decisions about theories are caused by social factors and processes, rather than by reasoning and logic. David Bloor is a sociologist of science and one of the founding members of the Science Studies Unit at Edinburgh University. Along with his colleagues, Barry Barnes and David Edge, Bloor is a leading proponent of the so-called strong programme in the sociology of knowledge. The core tenets of the strong programme are described and defended in his book *Knowledge and Social Imagery*.³⁵ Mary Hesse is a distinguished philosopher of science who taught for many years in the History and Philosophy of Science Programme at Cambridge University. Although not herself an advocate of Bloor's version of the sociology of science, Hesse is sympathetic to certain aspects of the strong programme and her writings are often cited favorably by Bloor. Laudan is particularly concerned with Hesse's paper "The Strong Thesis of Sociology of Science," which appeared as the second chapter of her book, *Revolutions and Reconstructions in the Philosophy of Science*.³⁶

Elsewhere, Laudan has vigorously attacked the strong programme, accusing it of error and confusion.³⁷ Since the issue before us now is not the truth or plausibility of the strong programme but merely the merits of two arguments (Hesse's and Bloor's) given in its defense, we need not discuss the strong programme itself in detail. A few brief comments will suffice before we address the specific arguments that Laudan attacks.

According to David Bloor, the defining characteristics of the strong programme are causality, impartiality, symmetry, and reflexivity.³⁸ Bloor claims that each of these conditions reflect procedural assumptions common to the established sciences. Only by adopting them, Bloor argues, can the study of scientific beliefs itself become "scientific."

The causal condition simply means that sociologists should identify the conditions that bring about scientific belief or states of knowledge, paying particular attention to those cultural and social features often presumed (falsely, in Bloor's view) to lie outside of science proper and the mechanisms within science by which belief is produced.³⁹

Impartiality is the requirement that explanations of scientific beliefs should be "impartial with respect to truth and falsity, rationality and irrationality, success or failure." Bloor insists that a scientific belief's being

true or rational does not, by itself, explain why a particular scientist holds that belief, and Bloor often repeats that rational beliefs stand just as much in need of explanation as do irrational ones. Thus, Bloor rejects the views held by philosophers such as R. G. Collingwood and Imre Lakatos that: (1) an action or belief can be explained only if it can be shown to be rational (Collingwood); (2) when we have shown that a person's action or belief is rational, no further explanation is needed or possible—the explanation of a human belief or action stops with the exhibition of its rationality (Collingwood and Lakatos); and (3) explanations in terms of sociological and psychological causes are appropriate only when the action or belief is irrational (Lakatos).⁴⁰ Since much of science is presumed to be rational, if (3) were true, very little would be left for the sociologist of science to do. Indeed, until the advent of the strong programme, the view that, because they are rationally well founded, most scientific beliefs are off-limits to sociology had been accepted by most sociologists of science: sociology may have much to tell us about the structure and organization of science, but it can do little or nothing to explain the content of scientific belief. Bloor denies this, claiming all of science, rational and irrational alike, as an appropriate subject of sociological explanation. Arguably, Bloor's second condition (impartiality) follows from his first condition (causality). If all explanation is causal and all beliefs (whether true or false, rational or irrational) have causes, then all beliefs (whether true or false, rational or irrational) can be explained in terms of those causes. This, of course, leaves it completely open as to what kinds of cause (perception, experience, reasoning, psychological conditioning, social pressure) are causally involved in the production of any particular belief. Hence the importance of Bloor's third condition, symmetry.

According to Bloor, symmetry requires that all scientific beliefs should receive the same style of explanation and that "the same type of cause would explain, say, true and false beliefs." Obviously, much depends on what Bloor means here by "the same type of cause." As Laudan has argued elsewhere, to claim that all beliefs, rational and irrational alike, are produced by the *same* causes—whether they be what the scientist takes to be good evidence, a neurophysiological mechanism, a psychological compulsion, education, peer pressure, political ideology, religion, class, or wealth—seems quite implausible. Contrary to Bloor's claim, symmetry is not a condition common to the well-established sciences. For example, nowhere in physics, chemistry, or geology do we find it stipulated in advance of empirical investigation that some broad class of phenomena must be the result of the same kind of cause in order for theories that attempt to explain those phenomena to qualify as scientific.⁴¹

The fourth and final condition, reflexivity, states that the patterns of explanation used by sociologists should be applicable to their own discipline. Bloor thinks that this condition must be imposed "because otherwise sociology would be a standing refutation of its own theories."

With some idea in hand of what the strong programme amounts to, let us consider Hesse and Bloor's arguments in favor of the strong programme's approach to explaining scientific belief and Laudan's criticism of those arguments. Hesse's argument begins from the fact of Humean underdetermination (HUD) and runs (with some embellishment) as follows:

- 1 HUD—scientific theories are deductively underdetermined by the data.
- 2 So, scientists must adopt extraempirical criteria for what counts as a good theory when deciding to accept one theory in preference to its empirically adequate rivals.
- 3 These extraempirical criteria differ over time and between groups.
- 4 Hence, the adoption of these criteria should be explained by social rather than logical factors.
- 5 Thus, the decision to accept particular scientific theories on the basis of these criteria must also be explained by social rather than logical factors.

Laudan has no quibble with premises (1) and (3) or the intermediate conclusion (2). His complaint is that (4)—the conclusion that the adoption of extraempirical criteria for evaluating theories must be socially caused—does not follow from them. He charges that Hesse is assuming, without argument, that anything that is not determined by deductive logic must be the product of social factors. But, he protests, why should we think that the decision to adopt a particular set of ampliative rules cannot be the result of reasoning? Only historical investigation can reveal what caused a particular scientist to adopt a particular set of ampliative rules. From the mere fact of deductive underdetermination of theories by data nothing at all follows about the nature of those causes: they may be sociological or they may not.

Although Laudan does not discuss it explicitly, he obviously thinks that premise (3)—the fact that different scientists in different periods have adopted different extraempirical criteria for choosing among theories—is irrelevant to Hesse's conclusion. The different values adopted, or the different weights attached to values such as simplicity, explanatory scope, and fertility, imply nothing about what caused individual scientists to adopt them. Indeed, in every case the adoption may have been the result of deductive reasoning, the difference in outcome being due solely to the different premises from which the scientists started. Another, more likely, possibility is that different scientists not only used different starting assumptions but also used both deductive and nondeductive reasoning to reach their conclusions about how rival scientific theories should be assessed.

Having dismissed Hesse's argument as a non sequitur, Laudan turns his attention to Bloor. The argument of Bloor's that Laudan criticizes is taken from a special issue of the journal *Studies in History and Philosophy of Science*.⁴² The lead article in that issue, a piece by Bloor, was followed by responses by a number of critics to whom Bloor, in turn, replied. One of those critics was Gerd Buchdahl (an editor of the journal), and Bloor's argument comes from his reply to Buchdahl's criticisms.⁴³ Buchdahl was especially critical of the following passage from Bloor's paper:

What is it that can then account for the known stability of our explicit theoretical knowledge? For the sociologist the answer is simple.

Such stability as there is in a system of knowledge comes entirely from the collective decisions of its creators and users. That is to say, from the requirement that certain laws and classifications be kept intact, and all adjustments and alterations carried out elsewhere We need not assume that a protected law or classification is singled out because of any intrinsic properties like truth, self-evidence or plausibility. Of course, such properties will be imputed to them, but this will be a justification for the special treatment rather than the cause of it.⁴⁴

Buchdahl took exception to Bloor's pronouncement that the "stability" of some laws in the sciences (that is, the fact these laws have been accepted by scientists over a long period of time) comes *entirely* from the decision of scientists (the law's "creators and users") to protect the law from refutation. To Buchdahl it seemed obvious that the truth of a law and the evidence for it can also have much to do with the law's stability. If the law were true, then the evidence would run in its favor; scientists would accept the law on the basis of that evidence; and thus the fact that the law is true would, contrary to Bloor, play a role in explaining why scientists have retained it. In his reply to Buchdahl on this point, Bloor writes:

If my use of the word "completely" gave the impression that sensory input had nothing whatever to do with the resulting system of knowledge, then it was indeed the wrong word. The fact is, however, that with no change whatsoever in their evidential basis, systems of belief can be and have been destabilized. Conversely they can be and have been held stable in the face of rapidly changing and highly problematic inputs from experience. So the stability of a system of a belief is the prerogative of its users.⁴⁵

It is this argument of Bloor's that Laudan criticizes in his paper. Laudan charges that Bloor's argument, like Hesse's argument discussed earlier, is a glaring non sequitur: its conclusion does not follow from its premises. Laudan accepts that the two premises are true: scientists have sometimes

changed their theoretical beliefs without that change being prompted by new evidence; scientists have also sometimes retained their theoretical beliefs despite new evidence that seemed to refute them. But from this it scarcely follows that "the stability of a belief is the prerogative of its users." The fact that in some cases evidence has not been decisive in causing scientists to change or retain their beliefs does not imply that no evidence can ever play such a role.

In judging Bloor's argument invalid, Laudan interprets Bloor's conclusion as asserting, in effect, that scientists can rationally hold onto any belief, regardless of the evidence against it. But in the final paragraph of his section on Bloor, Laudan interprets Bloor's conclusion as saying, instead, that no evidence can ever compel a rational scientist to change his beliefs. These two interpretations of Bloor's conclusion are not quite the same. For example, it might still be irrational for a scientist to retain a particular belief in the face of refuting evidence, even though that evidence does not compel rejection of the belief, either logically or causally. (Logical "compulsion" would presumably arise only if the evidence entailed that the belief is false, and, as Laudan keeps reminding us, there is more to rationality than deductive logic.) But the important point is Laudan's insistence that the belief change be a rational one. For, as he argued in evaluating Quine's version of the underdetermination thesis, without the restriction to the rational retention of belief, the contention that any belief can be retained "come what may" is trivial. Thus, Laudan concludes that Bloor's argument from underdetermination does nothing to make plausible the strong programme's claim that scientific beliefs are always caused by social forces and interests.

3.5 | Summary

When Pierre Duhem published his views on holism and the ambiguity of falsification at the turn of the century, he could not have guessed how they would be used and misused by later philosophers of science. Unlike the more recent doctrines that have flown under the banner of the Duhem-Quine thesis, Duhem's own version of holism is moderate, qualified, and plausible. Duhem argues that individual physical theories and postulates cannot be tested in isolation. Testable predictions (in the form of observation statements) can be made only when a single theory or postulate is combined with other theories, hypotheses, and assumptions. Observation or experiment may reveal that a prediction is false, but deductive logic (in the form of *modus tollens*) cannot tell us which theory, hypothesis, or assumption is at fault. All we can deduce is that at least one of these is false. To that extent, falsification is ambiguous.

Duhem was especially concerned to deny that there can be crucial

experiments in physics. Since a crucial experiment would not only falsify one theory but also establish its rival theory as true or highly probable, then clearly no experiment can be crucial if falsification is ambiguous. Moreover, even if an experiment could refute a particular theory, the rival theory need not be true or probable. In this respect the logic of testing in physics differs from the *reductio ad absurdum* reasoning used in logic and mathematics.

Duhem was equally critical of inductivism, arguing that no theory could be "deduced from the phenomena." Newton and Ampère had both claimed that each individual postulate of their theories could be inferred directly from observational facts or from experimental laws that, in turn, had been derived from such facts. Holism implies that this is impossible, since no single postulate by itself has any observational consequences. Duhem also pointed out that if Newton's gravitational theory is true, then the laws from which its postulates are allegedly deduced (Kepler's laws of planetary motion) must be false. Hence, again, inductivism fails.

Duhem was well aware that, because of the ambiguity of falsification, protecting a theory from refutation by making changes to the auxiliary hypotheses and assumptions used in testing that theory was possible, at least in principle. But he did not conclude from this that no theory can ever be falsified or that any scientific principle, however well entrenched, is immune to refutation. For example, Duhem criticized Poincaré's conventionalism, which held that many high-level principles in physics are not empirical hypotheses but definitions. Duhem agreed with Poincaré that individual principles (such as Newton's first law of motion) could not be tested in isolation. But he denied that it follows that these principles are nonempirical definitions. For, as with any other postulate in science, these principles can be tested indirectly, once they are combined with auxiliary hypotheses and assumptions. Thus, Duhem used his holist thesis to argue for the contention that all physical theories and postulates are testable and, in principle, falsifiable.

Duhem's account of how falsification can occur is rather sketchy and relies on what he calls scientific "good sense." Suppose there are two rival theories. One theory is fruitful in suggesting new experiments and its predictions prove successful. The other is less successful; with growing frequency, it and its associated auxiliaries have to be repaired and modified in ways that are often arbitrary, in order to protect the theory from refutation. Under these circumstances "good sense" tells the scientist that it is time to give up the ailing theory in favor of its more successful rival. Logic does not force the decision, but as time passes not adopting the more successful theory becomes increasingly unreasonable and irrational.

Quine was ignorant of Duhem's work when he wrote "Two Dogmas of Empiricism." Nonetheless, he used premises about holism and the ambiguity of falsification resembling Duhem's to reach controversial conclusions about meaning, analyticity, and a priori knowledge. In this paper,

he attacked two empiricist dogmas held by the logical positivists of the Vienna Circle—reductionism and analyticity. Reductionism, in its purest form, is the thesis that each meaningful synthetic statement is logically equivalent to (can be reduced to) the set of experiential statements (observation statements) that it implies. The experiential statements can be verified directly. A synthetic statement is conclusively verified when all of its experiential consequences are verified. According to the verifiability principle of meaning embraced by the logical positivists, the meaning of a synthetic statement is identified with the set of its verifiable consequences. So, for example, if two statements have exactly the same consequences, then they are synonymous. Synthetic statements that have no empirical consequences are judged to be meaningless (or, at least, to lack any cognitive meaning). Analyticity is the doctrine that certain statements having no empirical consequences (such as those of logic and mathematics) are true solely in virtue of the meanings of the words and terms used to express them. The logical positivists argued that it is because the propositions of logic and mathematics are analytic that we have certain *a priori* knowledge of their truth.

Quine rejects reductionism on the grounds that no single statement by itself logically implies anything about experience or observation. Statements have observationally verifiable consequences only when combined with other statements. With regard to science, Quine proclaims that the whole of science is needed to derive testable consequences. (In Quine's terminology, the smallest unit of empirical significance is science as a whole.) Thus, in "Two Dogmas," Quine espouses a global version of holism. He criticizes the doctrine of analyticity on the grounds that no satisfactory account has been given of the notion that a statement is true solely in virtue of its meaning. In particular, he rejects both Frege's characterization of analytic statements as those that reduce to tautologies by synonym substitution and Carnap's appeal to semantical rules in artificial languages. In light of these failures to explain analyticity, Quine suggests that there is no such thing. No statements are analytic, and none are *a priori*. In principle, any statement, even one from logic or mathematics, could be revised or abandoned in light of experience. Conversely, Quine also insists that any statement can be retained as true, regardless of experience, if we make drastic enough changes to the rest of our system of belief. Thus, he concludes by espousing a thesis of radical underdetermination: observation and evidence do not determine (or even constrain) which individual theories we decide to accept and which we choose to reject.

Gillies explores the similarities and differences between Duhem and Quine, criticizing aspects of each before offering a version of the Duhem-Quine thesis that he judges to be plausible. He rejects the global character of Quine's holism, faults Quine for paying little attention to rational constraints on theory choice besides deductive logic, and thinks that Quine

is wrong in claiming that no single scientific statement can be tested without the aid of auxiliary hypotheses and assumptions. He criticizes Duhem for not extending his holist thesis to sciences other than physics. In particular, Gillies sides with Quine by arguing (against Duhem) that mathematical theories and even logical principles (such as the law of excluded middle) could be given up as false in response to observation and experiment.

In "Demystifying Underdetermination" Laudan attacks a variety of arguments that have been used by philosophers and sociologists of science who are skeptical about the rationality and objectivity of science. Typically, these arguments appeal to some version of the underdetermination thesis in an attempt to show that there must be an important sociological, non-rational dimension to the decisions that scientists make about theories. In assessing these arguments, Laudan emphasizes the importance of distinguishing between deductive (Humean) and ampliative versions of the underdetermination thesis. According to the deductive underdetermination thesis, indefinitely many theories are logically consistent with any given body of evidence and (in conjunction with other statements) deductively entail that evidence. Ampliative underdetermination comes in many varieties. The most far-reaching is what Laudan calls the egalitarian thesis, namely, that every theory is as well supported by the evidence as any of its rivals. The thesis is ampliative because it talks about support, which includes such things as inductive confirmation, explanatory power, and simplicity. These go beyond the relations of logical consistency and deductive entailment to which the deductive thesis of underdetermination confines itself.

Laudan insists that, if it is not to be utterly trivial, Quine's assertion that any theory can be held as true regardless of the evidence must be understood normatively—as claiming that any theory can be held *rationally* whatever the evidence. He argues that deductive underdetermination is insufficient to support Quine's assertion and that Quine's assertion (with the crucial stipulation about rationality) presupposes the truth of the egalitarian thesis. For if the egalitarian thesis were false, then there would be times when the evidence would support one theory much better than its rivals, and, under those circumstances, it would be irrational to retain the inferior theory. Moreover, he contends that the further arguments Quine gives in "Two Dogmas" do nothing to show that any theory whatever can be rationally retained in the face of recalcitrant evidence. Laudan concludes that neither in "Two Dogmas" nor in his other writings has Quine made the egalitarian thesis plausible.

In the remainder of his article, Laudan looks at the work of Goodman, Kuhn, Hesse, and Bloor to see whether any of them has made a plausible case for ampliative underdetermination (in the form of the egalitarian thesis). Without venturing a solution to Goodman's new riddle of in-

duction, Laudan observes that even if it were true that some gruelike hypotheses are as well supported by our evidence as some nongruelike hypotheses, this hardly shows that any theory whatever enjoys the same support from that evidence. Kuhn defends a local version of the underdetermination thesis by restricting its scope to the known rivals of a given theory and arguing that shared methodological rules (involving values such as accuracy, scope, and simplicity) are too vague to pick out one theory as the best. Laudan counters by describing other rules, also widely shared by scientists, that are much less vague and which could, at least in principle, be used to resolve a contest between competing theories. Laudan illustrates his point with a historical example, showing how the Newtonians and the Cartesians settled their dispute about the shape of the earth by applying a simple methodological rule to the evidence they had gathered. Finally, Laudan examines the arguments of Hesse and Bloor in favor of the strong programme in the sociology of science. Laudan agrees with Hesse that deductive underdetermination implies that scientists must rely on extraempirical criteria in deciding among theories, but he rejects as gratuitous her implicit assumption that the adoption of such criteria must be caused by social factors. Similarly, Laudan finds no merit in Bloor's argument which, from premises about historical cases in which evidence has not determined the theoretical choices made by scientists, concludes that scientists can retain any theory or law regardless of any evidence. As in his criticism of Quine, so, too, in his response to Bloor, Laudan emphasizes the need to stipulate that it must be rational to retain the theory in the face of recalcitrant evidence if underdetermination is to be anything more than a boring triviality. Laudan concludes that none of the arguments he has examined do anything to make plausible an interesting version of the underdetermination thesis.

■ | Notes

- 1 In its most extreme form, inductivism claims that theories can be deduced (in some sense of *deduce*) from the observational phenomena. Duhem attacks this claim later in his article.
2. For a discussion of whether Duhem was correct in limiting his holist thesis to physics and excluding disciplines such as physiology and chemistry, see the later section, "Gillies' Criticisms of Duhem" (below).
3. Gillies, for example, thinks that it is false as a general claim about all theoretical hypotheses in science. See the section "Gillies' Version of the Duhem-Quine Thesis" (below) for an evaluation of this claim.
- 4 At the time of Foucault's experiment, the particle theory had already been largely abandoned in favor of the wave theory of Young and Fresnel. Whether

Foucault performed a crucial experiment is an epistemological judgment about the evidential significance of his work. It does not imply that Foucault's work was historically crucial in changing people's minds about the status of the rival theories (although it did have this effect on one important scientist, namely, Biot).

5. Duhem focuses exclusively on the conception of a crucial experiment as one having "the power to transform a physical hypothesis into an indisputable truth" (266), effectively ignoring the weaker alternative that a crucial experiment might merely establish a hypothesis as well confirmed. In our example, the rival theory T^* predicts the experimental observation $\sim O_1$ but, presumably, Duhem would regard this single successful prediction as insufficient to establish T^* as well-confirmed. Many philosophers of science would disagree with Duhem on this point, arguing that, in the right circumstances, a single successful prediction, especially a prediction of something considered highly improbable before the experiment was performed, could significantly raise the probability of T^* and thus make it well confirmed. See chapters 4 and 5 for a fuller discussion of this and related issues concerning confirmation.

6. For Popper's dismissal of induction and inductive confirmation, see the papers by Popper and Salmon in chapter 4 and the sections "Popper's Rejection of Induction" and "Salmon's Criticism of Popper's Anti-Inductivism" in the accompanying commentary.

7. Duhem does not consider trying to offer a more refined interpretation of Newton's method in terms of idealizations that give successively better approximations of the data. Newton's use of idealized evidence is explored in Ronald Laymon, "Newton's Demonstration of Universal Gravitation and Philosophical Theories of Confirmation," in *Testing Scientific Theories: Minnesota Studies in the Philosophy of Science*, vol. 10, ed. J. Earman (Minneapolis: University of Minnesota Press, 1983), 179–99.

8. In "Two Dogmas in Retrospect," *Canadian Journal of Philosophy* 21 (1991): 265–74, Quine confesses that, at the time he wrote his paper, he did not know about Duhem. It was Carl Hempel and Philipp Frank who brought Duhem to his attention.

9. In his later publications, Quine has drawn attention to the differences between his views and those of Duhem, distinguishing between Duhem's holism and his own far more sweeping thesis of empirical underdetermination. See, for example, W. V. Quine, "On Empirically Equivalent Systems of the World," *Erkenntnis* 9 (1975): 313–28.

10. This definition of analyticity is controversial. The characterization in terms of meanings is the one that Quine adopts, following Carnap. For some alternative definitions, see the next section, on the status of mathematical truths.

11. Thus, the relation between the two dogmas is, strictly speaking, neither one of identity nor even of entailment. In saying that the two dogmas are "identical," Quine is claiming merely that anyone who accepts one of the dogmas (reductionism) will have a strong psychological propensity to accept the other (analyticity) as well.

12. Carl G. Hempel, "On the Nature of Mathematical Truth," *American Mathematical Monthly* 52 (1945): 543–56.

13 Getting clear about the precise nature of the distinction between pure geometry and empirical geometry is no easy matter. For a defense of the distinction, see Carl G. Hempel, "Geometry and Empirical Science," *American Mathematical Monthly* 52 (1945): 7–17, and Ernest Nagel, *The Structure of Science* (London: Routledge & Kegan Paul, 1961), 215–76. For some of the difficulties with the distinction, see Alan Hausman, "Non-Euclidean Geometry and Relative Consistency Proofs," in *Motion and Time, Space and Matter*, ed. P. K. Machamer and R. G. Turnbull (Columbus: Ohio State University Press, 1976), 418–35.

14 Quine's remarks about the limitations of Carnap's characterization of analytic statements as those that come out true in all state-descriptions can be appreciated by thinking of truth tables. Consider Quine's own example of the compound statement, "John is a bachelor, and John is married." Obviously, this is a conjunction of the two statements $A =$ "John is a bachelor," and $B =$ "John is married." If A and B are regarded as atomic statements (and thus as ingredients in a state-description of a possible world), then one possible assignment of truth values to these atomic statements would have them both true. Thus, there would be at least one state-description (at least one possible world) in which "John is a bachelor, and John is married" is true. Similarly, there would be at least one other state-description (at least one other possible world) in which "John is a bachelor, and John is married" is false. Thus, the compound statement would be misclassified as synthetic. The problem, as Carnap was clearly aware, is that before employing his definition of analyticity, we first must ensure that all the statements that we take as atomic are logically independent of one another. In particular, we must rule out as candidates for atomicity any pair of sentences that contain synonyms or antonyms such as "bachelor" and "married man." Thus, the problem with this approach is the same as the problem with the other approaches Quine criticizes, namely, it fails to give a satisfactory account of synonymy.

15 As many of Quine's critics—notably H. P. Grice and P. F. Strawson—have pointed out, it is quite common for there to be a circle of expressions such that if any one member of the circle is understood or explained, then any other member of the circle can be understood or explained in terms of it. The terms *morally wrong*, *blameworthy*, and *violation of moral rules* belong to one such family circle, and the terms *analytic*, *synonymous*, *necessary*, and *semantical rule* belong to another. In arguing, as Quine does, that we cannot make "satisfactory sense" of analyticity, he seems to be assuming that a satisfactory account of an expression must give necessary and sufficient conditions for that expression without using any other member of the family circle to which it belongs. As a necessary condition for an expression's making sense, Quine's assumptions appear to be unreasonable, for there are many expressions that cannot be formally defined except in terms of members from the same family group. See H. P. Grice and P. F. Strawson, "In Defense of a Dogma," *Philosophical Review* 65 (1956): 141–58.

16 This is not the same as attributing to Quine the following, invalid, argument: if mathematical truths are analytic, then they are a priori, they are not analytic; therefore, they are not a priori. Rather, Quine's remarks about the revisability of mathematics and logic in the final section of his paper are intended to show that, contrary to popular belief, mathematics and logic are not a priori and, hence, they are not analytic.

17. Understandably, it is Quine's claim about the possibility of retaining a sentence "come what may" that has excited the interest of philosophers of science. But given Quine's attack on a priori knowledge (via his repudiation of the analytic-synthetic distinction), his claim about falsifying logical and mathematical principles is at least as radical and closer to Quine's real concern in "Two Dogmas." From Quine's perspective, (i) and (ii) are opposite sides of the same coin, since each follows from (a) Quine's doctrine that there is no difference in kind, either semantic or epistemological, between the two types of sentence traditionally classified as analytic and synthetic, respectively; (b) Quine's global version of holism; and (c) Quine's assumptions about the limitations of scientific method in deciding matters of confirmation and disconfirmation. From a more traditional perspective, (i) and (ii) seem curiously at odds with one another, for (i) seems to assert that any sentence could, in principle, be "made" a priori by the decision never to give it up, while (ii) asserts that no sentence is a priori in the sense of being immune to revision.

18. W. V. Quine, "On Empirically Equivalent Systems of the World," 315. Quine thinks that we often fail to recognize the unity that logic and mathematics confer on the sciences because, in the grip of the analytic-synthetic distinction, we place logic and mathematics in a different category from the rest of science.

19. W. V. Quine, *Philosophy of Logic* (Englewood Cliffs, N.J.: Prentice-Hall, 1970), 84.

20. One can appreciate even more vividly the failure of distribution in quantum logic by considering the following illustration. Let the variables X and P be the position and momentum of a simple system consisting of a single particle. Let X_1, X_2, \dots, X_n be a complete list of all the possible values of position that the particle can have, and let P_1, P_2, \dots, P_m be a complete list of all the possible values of its momentum. In quantum logic, the conjunction $(X_1 \vee X_2 \vee \dots \vee X_n) \& (P_1 \vee P_2 \vee \dots \vee P_m)$ is a necessary truth, a tautology. But the disjunction $(X_1 \& P_1) \vee \dots \vee (X_n \& P_1) \vee \dots \vee (X_n \& P_m)$ is a quantum logical contradiction. See Peter Gibbins, *Particles and Paradoxes: The Limits of Quantum Logic* (Cambridge: Cambridge University Press, 1987), 152.

21. One philosopher who has is Hilary Putnam. See his "Is Logic Empirical?" *Boston Studies in the Philosophy of Science*, vol. 5, ed. R. S. Cohen and M. W. Wartofsky (Dordrecht, Netherlands: D. Reidel, 1968), 216-41; revised version reprinted as "The Logic of Quantum Mechanics," *Mathematics, Matter, and Method, Philosophical Papers*, vol. 1, Cambridge: Cambridge University Press, 1975), 174-97.

22. But it is not inconsistent with the majority of "Two Dogmas," which attacks the notion that some statements are analytic because they can be reduced to tautologies by definitional substitution. One could deny that there are any statements that are analytic in this sense, while still affirming that the tautologies themselves are analytic, a priori truths.

23. For a dissenting view, see Hilary Putnam, "Is Logic Empirical?" and the discussion of Putnam's article in Peter Gibbins, *Particles and Paradoxes*, ch. 10. For Putnam, all the laws of the *true* logic are analytic, but we discover which they are empirically, by doing physics. Thus Putnam denies that all analytic truths are a

priori. He thinks that the logical connectives in quantum logic are the same as those in classical logic because they satisfy many of the same basic principles. What quantum mechanics teaches us, according to Putnam, is that the classical distribution laws governing *and*, *not*, and *or* are not analytic.

24 See Hilary Putnam, "There Is At Least One A Priori Truth," *Erkenntnis* 13 (1978): 153–70, reprinted in H. Putnam, *Realism and Reason* (New York: Cambridge University Press, 1983), 98–114. Putnam actually defends the a priori status, not of the classical law of noncontradiction (that no statement is both true and false) but of the weaker assertion that not every statement is both true and false (which, unlike the law of noncontradiction, allows for the possibility that some statements might have both truth values).

25 One feature of the organization of Laudan's article might confuse the reader. The first two sections after the introduction are labeled "Vintage Versions of Underdetermination" and "Ampliative Underdetermination," respectively. But the "Vintage" section includes discussions of both Hume and Quine, even though Quine is concerned largely, but not exclusively, with ampliative underdetermination. But there is a rationale for Laudan's inclusion of Quine in the same section as Hume: although Quine's versions of the underdetermination thesis are all ampliative, he never seriously discusses ampliative inference but confines himself almost entirely to the deductive relations of entailment and consistency. Indeed, Laudan argues that Quine comes perilously close to affirming doctrines of ampliative underdetermination solely on the basis of deductive underdetermination.

26 Although, of course, if Quine's (0) really does entail EGAL and we can show that EGAL is false, then we would have refuted (0). Laudan's aim in this paper is more modest, namely, to show that Quine has not given us any good reason to believe that (0) is true because he has failed to give us any good reason for accepting EGAL.

27 And indeed, as Laudan points out in a footnote, Quine himself conceded this in a letter to Adolf Grünbaum (dated 1 June 1962) published in Sandra Harding, ed., *Can Theories Be Refuted?* (Dordrecht, Netherlands: D. Reidel, 1976), 132. Quine admits that his thesis in "Two Dogmas"—the thesis Laudan has labelled QUD—is "probably trivial" and that his only purpose in advancing it in that paper was to illustrate his doctrine of semantic holism, that individual sentences do not have empirical content except as members of much larger groups of beliefs.

28 There are many definitions of *grue* in the literature spawned by Goodman's book. One popular variant drops the reference to "being examined" and the assumption about the persistence of the color of objects over time that are features of Goodman's original definition. Thus, it is quite common to find *grue* defined in the following way: *x* is *grue* at time *t* if and only if either *t* is earlier than *T* and *x* is green or *t* is not earlier than *T* and *x* is blue. See, for example, the definition of *grue* in Stephen F. Barker and Peter Achinstein, "On the New Riddle of Induction," *Philosophical Review* 69 (1960): 511–22. For a comprehensive annotated bibliography of writings about *grue*, see Douglas Stalker, ed., *Grue! The New Riddle of Induction* (La Salle, Ill.: Open Court, 1994). As Stalker notes, the word *gruebleen* was coined by James Joyce on page 23 of *Finnegan's Wake* (London: Faber and Faber, 1939), and Goodman had introduced a *grue*-like predicate as early as 1946, but the term *grue* itself was first used by Goodman in *Fact*,

Fiction, and Forecast (Cambridge, Mass.: Harvard University Press, 1955), 74. Since then, to quote Stalker, "the new riddle of induction has become a well-known topic in contemporary analytic philosophy—so well-known that only a philosophical hermit wouldn't recognize the word *grue*" (p. 2).

29 See Nelson Goodman, *Fact, Fiction, and Forecast*, ch. 4. Goodman's entrenchment theory is criticized in Rod J. Bertolet, "On the Merits of Entrenchment," *Analysis* 37 (1976): 29–31, and in Andrzej Zabłudowski, "Concerning a Fiction about How Facts Are Forecast," *Journal of Philosophy* 71 (1974): 97–112, reprinted in Stalker.

30 See Stephen F. Barker and Peter Achinstein, "On the New Riddle of Induction," *Philosophical Review* 69 (1960): 511–22, and Goodman's terse reply in "Positionality and Pictures" *Philosophical Review* 69 (1960): 523–25.

31. Popper denies that generalizations are ever confirmed by their instances because he rejects the whole notion of confirmation and inductive inference. See the section "Why All Theories Are Improbable" in the commentary on chapter 1, the papers by Popper and Salmon in chapter 4, and the sections "Popper's Rejection of Induction" and "Salmon's Criticism of Popper's Anti-inductivism" in the commentary on chapter 4. Other authors accept the notion of inductive confirmation but argue that generalizations are not always confirmed by their instances. See, for example, L. Jonathan Cohen, *The Implications of Induction* (London: Methuen, 1970), and Roger D. Rosenkrantz, "Does the Philosophy of Induction Rest on a Mistake?" *Journal of Philosophy* 79 (1982): 78–97.

32. See, for example, Paul Horwich, *Probability and Evidence* (Cambridge: Cambridge University Press, 1982), and Colin Howson and Peter Urbach, *Scientific Reasoning: The Bayesian Approach*, 2d ed. (La Salle, Ill.: Open Court, 1993).

33. Under the rubric of consistency, Kuhn includes both what Laudan calls internal consistency (freedom from logical contradiction) and what Kuhn calls external consistency (compatibility with other currently accepted theories). So another way of making Laudan's point is to say that one important component of what Kuhn calls consistency (namely, internal consistency) is not at all vague.

34. Further confirmation came with the invention of an accurate pendulum clock that was robust enough to be taken to sea. After the necessary corrections for variations in temperature, it was discovered that the clocks run slow when they are carried to the equator. This is exactly what Newton's theory predicts: because of the equatorial bulge, the acceleration due to gravity, g , is slightly less at the equator than it is further north or south. The periodic time of a simple pendulum of length L is given by $T = 2\pi\sqrt{L/g}$. As g decreases, T increases, and the clock runs slow.

35. David Bloor, *Knowledge and Social Imagery*, 2d ed. (Chicago, Ill.: University of Chicago, 1991). In the second edition of his book, Bloor weakened some of the claims attributed to the strong programme in the first edition. For a discussion of these changes and their significance, see Peter Slezak, "The Social Construction of Social Constructionism," *Inquiry* 37 (1994): 139–57. (The phrases *social constructionism*, *sociology of scientific knowledge*, and *social construction of knowledge* are often used to refer to the general movement to which the strong programme

belongs.) For appraisals of the strong programme from standpoints different from those of Laudan and Slezak, see Thomas McCarthy, "Scientific Rationality and the 'Strong Program' in the Sociology of Knowledge," in *Construction and Constraint*, ed. E. McMullin (Notre Dame, Ind.: University of Notre Dame Press, 1988), 73–96, and Arthur Fine, "Science Made Up: Constructivist Sociology of Scientific Knowledge," in *The Disunity of Science*, ed. P. Galison and D. J. Stump (Stanford, Calif.: Stanford University Press, 1996), 231–54.

36. Hesse's conclusion about the strong thesis (as she calls it) is guarded. According to Hesse, the strong thesis does not imply that there is no difference between rational rules and mere social conventions, nor does it entail that there are laws of scientific and social development. "All that is implied is the possibility of finding some correlations, amounting to historical explanations in particular cases, between types of scientific theory and particular social provenance. It may be felt that the 'strong' thesis has now become so weak as to be indistinguishable from something any rationalist or realist could accept in regard to the development of science." Mary B. Hesse, "The Strong Thesis of Sociology of Science," in *Revolutions and Reconstructions in the Philosophy of Science*, Bloomington: Indiana University Press, 1980), 56–57.

37 For Laudan's attack on the strong programme, see Larry Laudan, "The Pseudo-Science of Science?" *Philosophy of the Social Sciences* 11 (1981): 173–98; David Bloor responded to Laudan in "The Strengths of the Strong Programme," *Philosophy of the Social Sciences* 11 (1981): 199–213, finally, Larry Laudan published "More on Bloor," *Philosophy of the Social Sciences* 12 (1982): 71–74.

38 This list of characteristics is given in David Bloor, *Knowledge and Social Imagery*, p. 7. All the quotations attributed to Bloor are from this page of his book.

39. Among the latter, Bloor is a champion of what he calls *negotiation*. Negotiation, as Bloor describes it, is the informal, interpretative process by which logic and formal principles are applied to particular cases. Bloor thinks that the presentation of our reasoning in the form of deductive arguments is often an after-the-fact attempt to justify conclusions already reached on other grounds. Negotiation can also lead us to reject formal principles or even propositions in logic and mathematics that we formerly believed to be self-evident (such as the principle that the whole is always greater than the part, the denial of which can now serve as a definition of an infinite set). As far as Bloor is concerned, in logic and mathematics "there are no foundations other than social ones" (153). Here, as elsewhere in Bloor's manifesto for the strong programme, what professes to be simply the result of applying "the scientific method" to science itself actually rests on controversial claims about human psychology and a philosophical theory about a priori knowledge. For Bloor's account of negotiation, see David Bloor, *Knowledge and Social Imagery*, 131–56.

40 See R. G. Collingwood, *The Idea of History* (Oxford: Clarendon Press, 1946), and Imre Lakatos, "History of Science and Its Rational Reconstructions," in *Method and Appraisal in the Physical Sciences*, ed. C. Howson (Cambridge: Cambridge University Press, 1976), 1–39.

41. For these and other criticisms of the symmetry condition, see Larry Laudan, "The Pseudo-Science of Science?"

42. David Bloor, "Durkheim and Mauss Revisited. Classification and the Sociology of Knowledge," *Studies in History and Philosophy of Science* 13 (1982): 267-97.
43. Gerd Buchdahl, "Editorial Response to David Bloor," *Studies in History and Philosophy of Science* 13 (1982): 299-304, and David Bloor, "A Reply to Gerd Buchdahl," *Studies in History and Philosophy of Science* 13 (1982): 305-11.
44. David Bloor, "Durkheim and Mauss Revisited," 279-80.
45. David Bloor, "A Reply to Gerd Buchdahl," 306.



4 | Induction, Prediction, and Evidence

INTRODUCTION

Scientific laws and theories are universal generalizations that far outstrip the finite number of observations and experiments on which they are based. So it has seemed inevitable to philosophers of science from Aristotle to Carnap that some form of nondeductive inference must connect theories and laws with our evidence for them in some rational, justified way. Nearly always, induction and inductive inference are picked for this role. Scientific theories are confirmed, inductively, by evidence, and this inductive connection with evidence is, for the most part, what makes scientific theories reliable and our belief in them justified. Without inductive confirmation, science would be no better epistemically than blind guesswork, soothsaying, or wishful thinking.

Traditionally, the two main roles claimed for induction have been creative inference, which leads from evidence to the formulation of new theories (a logic of discovery) and confirmation, which connects evidence to theories after they have been formulated (a logic of justification).¹ These two roles have sometimes been connected by arguing that a theory generated in the right way—by inductive generalization from data—will automatically be justified without needing any new evidence to support it. This view has fallen into disfavor,² mainly because many theories in modern science appeal to entities and processes that are not, in any straightforward sense, observable.³ Theories postulate genes, quarks, and electrons, for example, but it would be impossible to infer these theories inductively from observation reports in which such terms are entirely absent. Moreover, the mathematical complexity of modern scientific theories makes it hard to imagine how they could possibly be generated by inferences as simple as inductive generalization. So induction as a logic of discovery

(and a fortiori as a discovery mechanism that brings justification with it) has been largely abandoned.

Most modern philosophers of science view induction exclusively as a logic of justification, to be applied to theories only after they have been generated. The most popular position is called the *hypothetico-deductive model*. On this model, a theory's origin—the way it was generated—is entirely irrelevant to its epistemic justification. All that matters is how the theory stands with respect to the things it explains and predicts. Many hypothetico-deductivists place a special emphasis on prediction, arguing that a theory is most powerfully confirmed when it successfully forecasts the outcome of new experiments and observations. Some of these philosophers go further, denying that the explanation of results already known has any power to confirm a theory. These philosophers, called *predictionists*, argue that only novel predictions count, inductively, as evidence. This debate within the ranks of the hypothetico-deductivists, between predictionism and more accommodating views of what can count as evidence, is one of several issues explored in this chapter.

Despite its apparent ubiquity in science, inductive reasoning is philosophically controversial. In "Induction," Peter Lipton explains why. He distinguishes between the problem of describing the general principles we follow when we reason inductively and the problem of showing that those inferences are justified. The second problem, often simply referred to as *the problem of induction*, originates with the skeptical argument of David Hume (1711–76), who notoriously concluded that it is impossible to show that induction is either reliable or reasonable. Karl Popper thinks that Hume was right about this: there is no justification of induction. In "The Problem of Induction," Popper rejects all inductive reasoning and, with it, the whole notion of inductive confirmation. In its place he offers a theory of scientific method—falsificationism—that relies solely on deduction. Wesley Salmon disputes the adequacy of Popper's falsificationism as an account of science. In "Rational Prediction" Salmon argues that, by dispensing with induction and inductive confirmation, Popper cannot justify our preference for using our best-tested theories when we need to make practical decisions about the future.

With Carl Hempel's "Criteria of Confirmation and Acceptability," the focus shifts away from the problem of justifying induction and towards articulating and assessing the various criteria that scientists use in weighing the merits of theories. These criteria include the quantity, diversity, and precision of evidence, as well as a theory's simplicity, its support from other (well-established) theories, and its ability to make novel predictions. Hempel pays particular attention to whether the value we place on simplicity is merely a subjective or pragmatic preference or whether it can be objectively justified in terms of inductive probability or greater falsifiability.

The last two readings examine in detail arguments for and against the importance of novel predictions in determining confirmation. In "Is Evi-

dence Historical?” Laura Snyder attacks the historical thesis according to which the power of a result to confirm a theory depends on when it was discovered. Predictionists insist that a result can confirm a theory only if it is discovered after the theory has been proposed. Explanationists assert exactly the opposite, arguing that only the explanation of previously known facts has the power to confirm a theory. Snyder thinks that both predictionists and explanationists are wrong and tries to show this by attacking the historical thesis to which both are committed. Her argument depends on accepting an objective characterization of evidence in science. Peter Achinstein agrees with Snyder that the historical thesis is false and in “Explanation v. Prediction: Which Carries More Weight?” concurs with Snyder that the distinction between explanation and prediction is not, by itself, relevant to confirmation. Nonetheless, he argues, historical information does sometimes determine whether a result is evidence for a theory and how strongly it confirms the theory. If, for example, an experiment is performed, then regardless of whether the outcome verifies a novel prediction or describes something previously known, how the experiment was conducted can affect confirmation.

■ | Notes

1. See the section “Why There Is No Logic of Discovery” in the commentary following the readings in this chapter.
2. For two contrasting accounts of where the failure of inductivism as an account of theory generation leaves the whole question of the relevance of generation to justification, see Thomas Nickles, “Justification and Experiment,” in *The Uses of Experiment*, ed. David Gooding, Trevor Pinch, and Simon Schaffer (Cambridge: Cambridge University Press, 1989), 299–333, and Larry Laudan, “Why Was the Logic of Discovery Abandoned?” in *Scientific Discovery, Logic, and Rationality*, ed. Thomas Nickles (Dordrecht, Netherlands: D. Reidel, 1980), 173–83.
3. The concept of observability is explored at length in chapter 9. See especially the first three papers in that chapter by Maxwell, Van Fraassen, and Musgrave and the discussion of them in the accompanying commentary.

Induction

■ | Underdetermination

Inductive inference is a matter of weighing evidence and judging likelihood, not of proof. How do we go about making these non-demonstrative judgments, and why should we believe they are reliable? Both the question of description and the question of justification arise from underdetermination. To say that an outcome is underdetermined is to say that some information about initial conditions and rules or principles does not guarantee a unique solution. The information that Tom spent five dollars on apples and oranges and that apples are fifty cents a pound and oranges a dollar a pound underdetermines how much fruit Tom bought, given only the rules of deduction. Similarly, those rules and a finite number of points on a curve underdetermine the curve, since there are many curves that would pass through those points.

Underdetermination may also arise in our description of the way a person learns or makes inferences. A description of the evidence, along with a certain set of rules, not necessarily just those of deduction, may underdetermine what is learned or inferred. Insofar as we have described all the evidence and the person is not behaving erratically, this shows that there are hidden rules. We can then study the patterns of learning or inference to try to discover them. Noam Chomsky's argument from 'the poverty of the stimulus' is a good example of how underdetermination can be used to disclose the existence of additional rules (1965, ch. 1, sec. 8, esp. pp. 58–9). Children learn the language of their elders, an ability that enables them to understand an indefinite number of sentences on first acquaintance. The talk young children hear, however, along with rules of deduction and any plausible general rules of induction, grossly underde-

termine the language they learn. What they hear is limited and often includes many ungrammatical sentences, and the little they hear that is well formed is compatible with many possible languages other than the one they learn. Therefore, Chomsky argues, in addition to any general principles of deduction and induction, children must be born with strong linguistic rules or principles that further restrict the class of languages they will learn, so that the actual words they hear are now sufficient to determine a unique language. Moreover, since a child will learn whatever language he is brought up in, these principles cannot be peculiar to a particular human language; instead, they must specify something that is common to all of them. For Chomsky, determining the structure of these universal principles and the way they work is the central task of modern linguistics.

Thomas Kuhn provides another well-known example of using underdetermination as a tool to investigate cognitive principles. He begins from an argument about scientific research strikingly similar to Chomsky's argument about language acquisition (1970, 1977, esp. ch. 12). In most periods in the history of a developed scientific specialty, scientists are in broad agreement about which problems to work on, how to attack them, and what counts as solving them. But the explicit beliefs and rules scientists share, especially their theories, data, general rules of deduction and induction, and any explicit methodological rules, underdetermine these shared judgments. Many possible judgments are compatible with these beliefs and rules other than the ones the scientists make. So Kuhn argues that there must be additional field-specific principles that determine the actual judgments. Unlike Chomsky, Kuhn does not argue for principles that are either innate or in the form of rules, narrowly construed. Instead, scientists acquire through their education a stock of exemplars—concrete problem solutions in their specialty—and use them to guide their research. They pick new problems that look similar to an exemplar problem, they try techniques that are similar to those that worked in that exemplar, and they assess their success by reference to the standards of solution that the exemplars illustrate. Thus the exemplars set up a web of 'perceived similarity relations' that guide future research, and the shared judgments are explained by the shared exemplars. These similarities are not created or governed by rules, but they result in a pattern of research that roughly mimics one that is rule-governed. Just how exemplars do this work, and what happens when they stop working, provide the focus of Kuhn's account of science.

Chomsky and Kuhn are both arguing for unacknowledged principles of induction. In these cases, however, underdetermination is taken to be a symptom of the existence of highly specialized principles, whether of language acquisition or of scientific method in a particular field, since the underdetermination is claimed to remain even if we include general principles of induction among our rules. But the same pattern of argument

applies to the general case. If an inference is inductive, then by definition it is underdetermined by the evidence and the rules of deduction. Insofar as our inductive practices are methodical, we must use additional rules or principles of inference, and we may study the patterns of our inferences in an attempt to discover them.

■ | Justification

The two central questions about our general principles of induction concern description and justification. What principles do we use, and how can they be shown to be good principles? The question of description seems at first to take priority. How can we even attempt to justify our principles until we know what they are? Historically, however, the justification question came first. One reason for this is that this question gets its grip from skeptical arguments that seem to apply to any principles that could account for the way we fill the gap between evidence and inference.

The problem of justification is to show that our inferential methods are good methods. The natural way to understand this is in terms of truth. We want our methods of inference to be 'truth-tropic', to take us towards the truth. For deduction, a good argument is one that is valid, a perfect truth conduit, where it is impossible for there to be true premises but a false conclusion. The problem of justification here would be to show that arguments we judge valid are in fact so. For induction, perfect reliability is out of the question. By definition, even a good inductive argument is one where it is possible for there to be true premises but a false conclusion. Moreover, it is clear that the reasonable inductive inferences we make are not 100 per cent reliable even in this world, since they sometimes sadly take us from truth to falsehood. Nevertheless, it remains natural to construe the task of justification as that of showing truth-tropism. We would like to show that those inductive inferences we judge worth making are ones that usually take us from true premises to true conclusions.

A skeptical argument that makes the problem of justification pressing has two components, underdetermination and circularity. The first is an argument that the inferences in question are underdetermined, given only our premises and the rules of deduction; that the premises and those rules are compatible not just with the inferences we make, but also with other, incompatible inferences. This shows that the inferences we make really are inductive and, by showing that there are possible worlds where the principles we use take us from true premises to false conclusions, it also shows that there are worlds where our principles would fail us. Revealing this underdetermination, however, does not yet generate a skeptical argument, since we might have good reason to believe that the actual world is one where our principles are reliable. So the skeptical argument requires

a second component, an argument for circularity, which attempts to show that we cannot rule out the possibility of unreliability that underdetermination raises without employing the very principles that are under investigation, and so begging the question.

Although it is not traditionally seen as raising the problem of induction, Descartes' 'First Meditation' is a classic illustration of this technique. Descartes' goal is to cast doubt on the 'testimony of the senses', which leads us to infer that there is, say, a mountain in the distance because that is what it looks like. He begins by arguing that we ought not to trust the senses completely, since we know that they do sometimes mislead us, 'when it is a question of very small and distant things'. This argument relies on the underdetermination component, since it capitalizes on the fact that the way things appear does not entail the way they are, but it does not yet have the circle component. We can corroborate our inferences about small and distant things without circularity by taking a closer look (cf. Williams, 1978, pp. 51–2). But Descartes immediately moves on from the small and the distant to the large and near. No matter how clearly we seem to see something, it may only be a dream, or a misleading experience induced by an evil demon. These arguments describe possible situations where even the most obvious testimony of the senses is misleading. Moreover, unlike the worry about small and distant things, these arguments also have a circle component. There is apparently no way to test whether a demon is misleading us with a particular experience, since any test would itself rely on experiences that he might have induced. The senses may be liars and give us false testimony, and we should not find any comfort if they also report that they are telling us the truth.

The demon argument begins with the underdetermination of observational belief by observational experience, construes the missing principle of inference on the model of inference from testimony, and then suggests that the reliability of this principle could only be shown by assuming it. Perhaps one of the reasons Descartes' arguments are not traditionally seen as raising the problem of justifying induction is that his response to his own skepticism is to reject the underdetermination upon which it rests. Descartes argues that, since inferences from the senses must be inductive and so raise a skeptical problem, our knowledge must instead have a different sort of foundation for which the problem of underdetermination does not arise. The *cogito* and the principles of clearness and distinctness that it exemplifies are supposed to provide the non-inductive alternative. In other words, the moral he draws from underdetermination and circularity is not that our principles of induction require some different sort of defense or must be accepted without justification, but that we must use different premises and principles, for which the skeptical problem does not arise. For a skeptical argument about induction that does not lead to the rejection of induction, we must turn to its traditional home, in the arguments of David Hume.

Hume also begins with underdetermination, in this case that our observations do not entail our predictions (1777, sec. IV). He then suggests that the governing principle of all our inductive inferences is that nature is uniform, that the unobserved (but observable) world is much like what we have observed. The question of justification is then the question of showing that nature is indeed uniform. This cannot be deduced from what we have observed, since the claim of uniformity itself incorporates a massive prediction. But the only other way to argue for uniformity is to use an inductive argument, which would rely on the principle of uniformity, leaving the question begged. According to Hume, we are addicted to the practice of induction, but it is a practice that cannot be justified.

To illustrate the problem, suppose our fundamental principle of inductive inference is 'More of the Same'. We believe that strong inductive arguments are those whose conclusions predict the continuation of a pattern described in the premises. Applying this principle of conservative induction, we would infer that the sun will rise tomorrow, since it has always risen in the past; and we would judge worthless the argument that the sun will not rise tomorrow since it has always risen in the past. It is, however, easy to come up with a factitious principle to underwrite the latter argument. According to the principle of revolutionary induction, 'It's Time for a Change', and this sanctions the dark inference. Hume's argument is that we have no way to show that conservative induction, the principle he claims we actually use for our inferences, will do any better than intuitively wild principles like the principle of revolutionary induction. Of course conservative induction has had the more impressive track record. Most of the inferences from true premises that it has sanctioned have also had true conclusions. Revolutionary induction, by contrast, has been conspicuous in failure, or would have been, had anyone relied on it. The question of justification, however, does not ask which method of inference has been successful, it asks which one will be successful.

Still, the track record of conservative induction appears to be a reason to trust it. That record is imperfect (we are not aspiring to deduction), but very impressive, particularly as compared with revolutionary induction and its ilk. In short, induction will work because it has worked. This seems the only justification our inductive ways could ever have or require. Hume's disturbing observation was that this justification appears circular, no better than trying to convince someone that you are honest by saying that you are. Much as Descartes argued that we should not be moved if the senses give testimony on their own behalf, so Hume argued that we cannot appeal to the history of induction to certify induction. The trouble is that the argument that conservative inductions will work because they have worked is itself an induction. The past success is not supposed to prove future success, only make it very likely. But then we must decide which standards to use to evaluate this argument. It has the form 'More of the Same', so conservatives will give it high marks, but since its conclusion is just to

underwrite conservatism, this begs the question. If we apply the revolutionary principle, it counts as a very weak argument. Worse still, by revolutionary standards, conservative induction is likely to fail precisely because it has succeeded in the past, and the past failures of revolutionary induction augur well for its future success (cf. Skyrms, 1986, ch. 2). The justification of revolutionary induction seems no worse than the justification of conservative induction, which is to say that the justification of conservative induction looks very bad indeed.

The problem of justifying induction does not show that there are other inductive standards better than our own. Instead it argues for a deep symmetry: many sets of standards, most of them wildly different from our own and incompatible with each other, are yet completely on a par from a justificatory point of view. This is why the problem of justification can be posed before we have solved the problem of description. Whatever inductive principles we use, the fact that they are inductive seems enough for the skeptic to show that they defy justification. We fill the gap of underdetermination between observation and prediction in one way, but it could be filled in many other ways that would have led to entirely different predictions. We have no way of showing that our way is any better than any of these others. It is not merely that the revolutionaries will not be convinced by the justificatory arguments of the conservatives: the conservatives should not accept their own defense either, since among their standards is one which says that a circular argument is a bad argument, even if it is in one's own aid. Even if I am honest, I ought to admit that the fact that I say so ought not carry any weight. We have a psychological compulsion to favor our own inductive principles but, if Hume is right, we should see that we cannot even provide a cogent rationalization of our behavior.

It seems to me that we do not yet have a satisfying solution to Hume's challenge and that the prospects for one are bleak, but there are other problems of justification that are more tractable. The peculiar difficulty of meeting Hume's skeptical argument against induction is that he casts doubt on our inductive principles as a whole, and so any recourse to induction to justify induction seems hopeless. But one can also ask for the justification of particular inductive principles and, as Descartes' example of small and distant things suggests, this leaves open the possibility of appeal to other principles without begging the question. For example, among our principles of inference is one that makes us more likely to infer a theory if it is supported by a variety of evidence than if it is supported by a similar amount of homogeneous data. This is the sort of principle that might be justified in terms of a more basic inductive principle, say that we have better reason to infer a theory when all the reasonable competitors have been refuted, or that a theory is only worth inferring when each of its major components has been separately tested. Another, more controversial, example of a special principle that might be justified

without circularity is that, all else being equal, a theory deserves more credit from its successful predictions than it does from data that the theory was constructed to fit. This appears to be an inductive preference most of us have, but the case is controversial because it is not at all obvious that it is rational. On the one hand, many people feel that only a prediction can be a real test, since a theory cannot possibly be refuted by data it is built to accommodate; on the other, that logical relations between theory and data upon which inductive support exclusively depends cannot be affected by the merely historical fact that the data were available before or only after the theory was proposed. In any event, this is an issue of inductive principle that is susceptible to non-circular evaluation. . . . Finally, though a really satisfying solution to Hume's problem would have to be an argument for the reliability of our principles that had force against the inductive skeptic, there may be arguments for reliability that do not meet this condition yet still have probative value for those of us who already accept some forms of induction. . . .

■ | Description

We can now see why the problem of justification, the problem of showing that our inductive principles are reliable, did not have to wait for a detailed description of those principles. The problem of justifying our principles gets its bite from skeptical arguments, and these appear to depend only on the fact that these principles are principles of induction, not on the particular form they take. The crucial argument is that the only way to justify our principles would be to reason with the very same principles, which is illegitimate; an argument that seems to work whatever the details of our inferences. The irrelevance of the details comes out in the symmetry of Hume's argument: just as the future success of conservative induction gains no plausibility from its past success, so the future success of revolutionary induction gains nothing from its past failures. As the practice varies, so does the justificatory argument, preserving the pernicious circularity. Thus the question of justification has had a life of its own: it has not waited for a detailed description of the practice whose warrant it calls into doubt.

By the same token, the question of description has fortunately not waited for an answer to the skeptical arguments. Even if our inferences were unjustifiable, one still might be interested in saying how they work. The problem of description is not to show that our inferential practices are reliable; it is simply to describe them as they stand. One might have thought that this would be a relatively trivial problem. First of all, there are no powerful reasons for thinking that the problem is impossible, as there are for the problem of justification. There is no great skeptical ar-

gment against the possibility of description. It is true that any account of our principles will require inductive support, since we must see whether it jibes with our observed inductive practice. This, however, raises no general problem of circularity now that justification is not the issue. Using induction to investigate induction is no more a problem here than using observation to study the structure and function of the eye. Second, it is not just that a solution to the problem of describing our inductive principles should be possible, but that it should be easy. After all, they are our principles, and we use them constantly. It thus comes as something of a shock to discover how extraordinarily difficult the problem of description has turned out to be. It is not merely that ordinary reasoners are unable to describe what they are doing; at least one hundred and fifty years of focused effort by epistemologists and philosophers of science has yielded little better. Again, it is not merely that we have yet to capture all the details, but that the most popular accounts of the gross structure of induction are wildly at variance with our actual practice.

Why is description so hard? One reason is a quite general gap between what we can do and what we can describe. You may know how to do something without knowing how you do it; indeed, this is the usual situation. It is one thing to know how to tie one's shoes or to ride a bike, it is quite another thing to be able to give a principled description of what it is that one knows. Chomsky's work on principles of language acquisition and Kuhn's work on scientific method are good cognitive examples. Their investigations would not be so important and controversial if the ordinary speaker of a language knew how she distinguished grammatical from ungrammatical sentences or the normal scientist knew how he made his methodological judgments. The speaker and the scientist employ various principles, but they are not conscious of them. The situation is similar in the case of inductive inference generally. Although we may partially articulate some of our inferences if, for example, we are called upon to defend them, we are not conscious of the fundamental principles of inductive inference we constantly use.

Since our principles of induction are neither available to introspection, nor observable in any other way, the evidence for their structure must be highly indirect. The project of description is one of black box inference, where we try to reconstruct the underlying mechanism on the basis of the superficial patterns of evidence and inference we observe in ourselves. This is no trivial problem. Part of the difficulty is simply the fact of underdetermination. As the examples of Chomsky and Kuhn show, underdetermination can be a symptom of missing principles and a clue to their nature, but it is one that does not itself determine a unique answer. In other words, where the evidence and the rules of deduction underdetermine inference, that information also underdetermines the missing principles. There will always be many different possible mechanisms that would produce the same patterns, so how can one decide which one is

actually operating? In practice, however, we usually have the opposite problem: we can not come up with even one description that would yield the patterns we observe. The situation is the same in scientific theorizing generally. There is always more than one account of the unobserved and often unobservable world that would account for what we observe, but scientists' actual difficulty is often to come up with even one theory that fits the observed facts. On reflection, then, it should not surprise us that the problem of description has turned out to be so difficult. Why should we suppose that the project of describing our inductive principles is going to be easier than it would be, say, to give a detailed account of the working of a computer on the basis of the correlations between keys pressed and images on the screen?

Now that we are prepared for the worst, we may turn to some of the popular attempts at description. In my discussion of the problem of justification, I suggested that, following Hume's idea of induction as habit formation, we describe our pattern of inference as 'More of the Same'. This is pleasingly simple, but the conservative principle is at best a caricature of our actual practice. We sometimes do not infer that things will remain the same and we sometimes infer that things are going to change. When my mechanic tells me that my brakes are about to fail, I do not suppose that he is therefore a revolutionary inductivist. Again, we often make inductive inferences from something we observe to something invisible, such as from people's behavior to their beliefs or from the scientific evidence to unobservable entities and processes, and this does not fit into the conservative mold. 'More of the Same' might enable me to predict what you will do on the basis of what you have done (if you are a creature of habit), but it will not tell me what you are or will be thinking.

Faced with the difficulty of providing a general description, a reasonable strategy is to begin by trying to describe one part of our inductive practice. This is a risky procedure, since the part one picks may not really be describable in isolation, but there are sometimes reasons to believe that a particular part is independent enough to permit a useful separation. Chomsky must believe this about our principles of linguistic inference. Similarly, one might plausibly hold that, while simple habit formation cannot be the whole of our inductive practice, it is a core mechanism that can be treated in isolation. Thus one might try to salvage the intuition behind the conservative principle by giving a more precise account of the cases where we are willing to project a pattern into the future, leaving to one side the apparently more difficult problems of accounting for predictions of change and inferences to the unobservable. What we may call the instantial model of inductive confirmation may be seen in this spirit. According to it, a hypothesis of the form 'All A's are B' is supported by its positive instances, by observed A's that are also B (cf. Hempel, 1965, ch. 1). This is not, strictly speaking, an account of inductive *inference*,

since it does not say either how we come up with the hypothesis in the first place or how many supporting instances are required before we actually infer it, but this switching of the problem from inference to support may also be taken as a strategic simplification. In any event, the underlying idea is that, if enough positive instances and no refuting instances (A's that are not B) are observed, we will infer the hypothesis, from which we may then deduce the prediction that the next A we observe will be B.

This model could only be a very partial description of our inductive principles but, within its restricted range, it strikes many people initially as a truism, and one that captures Hume's point about our propensity to extend observed patterns. Observed positive instances are not necessary for inductive support, as inferences to the unobserved and to change show, but they might seem at least sufficient. The instantial model, however, has been shown to be wildly over-permissive. Some hypotheses are supported by their positive instances, but many are not. Observing only black ravens may lead one to believe that all ravens are black, but observing only bearded philosophers would probably not lead one to infer that all philosophers are bearded. Nelson Goodman has generalized this problem, by showing how the instantial model sanctions any prediction at all if there is no restriction on the hypotheses to which it can be applied (Goodman, 1983, ch. 3). His technique is to construct hypotheses with factitious predicates. * Black ravens provide no reason to believe that the next swan we see will be white, but they do provide positive instances of the artificial hypothesis that 'All raveswans are blight', where something is a raveswan just in case it is either observed before today and a raven, or not so observed and a swan, and where something is blight just in case it is either observed before today and black, or not so observed and white. But the hypothesis that all raveswans are blight entails that the next observed raveswan will be blight which, given the definitions, is just to say that the next swan will be white.

The other famous difficulty facing the instantial model arises for hypotheses that do seem to be supported by their instances. Black ravens support the hypothesis that all ravens are black. This hypothesis is logically equivalent to the contrapositive hypothesis that all non-black things are non-ravens: there is no possible situation where one hypothesis would be true but the other false. According to the instantial model, the contrapositive hypothesis is supported by non-black non-ravens, such as green leaves. The rub comes with the observation that whatever supports a hypothesis also supports anything logically equivalent to it. This is very plausible, since support provides a reason to believe true, and we know that if a hypothesis is true, then so must be anything logically equivalent to it. But

* See the section "Goodman's Gruesome New Riddle of Induction" in the commentary on chapter 3.

then the instantial model once again makes inductive support far too easy, counting green leaves as evidence that all ravens are black (Hempel, 1965, ch. 1).

Another account of inductive support is the hypothetico-deductive model (Hempel, 1966, chs 2, 3). On this view, a hypothesis or theory is supported when it, along with various other statements, deductively entails a datum. Thus a theory is supported by its successful predictions. This account has a number of attractions. First, although it leaves to one side the important question of the source of hypotheses, it has much wider scope than the instantial model, since it allows for the support of hypotheses that appeal to unobservable entities and processes. The big bang theory of the origin of the universe obviously cannot be directly supported but, along with other statements, it entails that we ought to find ourselves today traveling through a uniform background radiation, like the ripples left by a rock falling into a pond. The fact that we do now observe this radiation (or effects of it) provides some reason to believe the big bang theory. Thus, even if a hypothesis cannot be supported by its instances, because its instances are not observable, it can be supported by its observable logical consequences. Second, the model enables us to co-opt our accounts of deduction for an account of induction, an attractive possibility since our understanding of deductive principles is so much better than our understanding of inductive principles. Lastly, the hypothetico-deductive model seems genuinely to reflect scientific practice, which is perhaps why it has become the scientists' philosophy of science.

In spite of all its attractions, our criticism of the hypothetico-deductive model here can be brief, since it inherits all the over-permissiveness of the instantial model. Any case of support by positive instances will also be a case of support by consequences. The hypothesis that all A's are B, along with the premise that an individual is A, entails that it will also be B, so the thing observed to be B supports the hypothesis, according to the hypothetico-deductive model. That is, any case of instantial support is also a case of hypothetico-deductive support, so the model has to face the problem of unsupportable hypotheses and the raven paradox. Moreover, the hypothetico-deductive model is similarly over-permissive in the case of vertical inferences to hypotheses about unobservables, a problem that the instantial model avoided by ignoring such inferences altogether. The difficulty is structurally similar to Goodman's problem of fictitious predicates. Consider the conjunction of the hypotheses that all ravens are black and that all swans are white. This conjunction, along with premises concerning the identity of various ravens, entails that they will be black. According to the model, the conjunction is supported by black ravens, and it entails its own conjunct about swans. The model thus appears to sanction the inference from black ravens to white swans (cf. Goodman, 1983, pp. 67–8). Similarly, the hypothesis that all swans are white taken alone entails the inclusive disjunction that either all swans are white or there is

a black raven, a disjunction we could establish by seeing a black raven, thus giving illicit hypothetico-deductive support to the swan hypothesis. These maneuvers are obviously artificial, but nobody has managed to show how the model can be modified to avoid them without also eliminating most genuine cases of inductive support (cf. Glymour, 1980, ch. 2). Finally, in addition to being too permissive, finding support where none exists, the model is also too strict, since data may support a hypothesis which does not, along with reasonable auxiliary premises, entail them.

We have now canvassed three attempts to tackle the descriptive problem 'More of the Same', the instantial model, and the hypothetico-deductive model. The first two could, at best, account for a special class of particularly simple inferences, and all three are massively over-permissive, finding inductive support where there is none to be had. They do not give enough structure to the black box of our inductive principles to determine the inferences we actually make. This is not to say that these accounts describe mechanisms that would yield too many inferences: they would probably yield too few. A 'hypothetico-deductive box', for example, would probably have little or no inferential output, given the plausible additional principle that we will not make inferences we know to be contradictory. For every hypothesis that we would be inclined to infer on the basis of the deductive support it enjoys, there will be an incompatible hypothesis that is similarly supported, and the result is no inference at all, so long as both hypotheses are considered and their incompatibility recognized.

A fourth account of induction, the last I will consider in this section, focuses on causal inference. It is a striking fact about our inductive practice, both lay and scientific, that so many of our inferences depend on inferring from effects to their probable causes. This is something that Hume himself emphasized (Hume, 1777, sec. IV). The cases of direct causal inference are legion, such as the doctor's inference from symptom to disease, the detective's inference from evidence to perpetrator, the mechanic's inference from the engine noises to what is broken, and many scientific inferences from data to theoretical explanation. Moreover, it is striking that we often make a causal inference even when our main interest is in prediction. Indeed, the detour through causal theory on the route from data to prediction seems to be at the center of many of the dramatic successes of scientific prediction. All this suggests that we might do well to consider an account of the way causal inference works as a central component of a description of our inductive practice.

The best known account of causal inference is John Stuart Mill's discussion of the 'methods of experimental inquiry' (Mill, 1904, book III, ch. VIII). The two central methods are the Method of Agreement and especially the Method of Difference. According to the Method of Agreement, in idealized form, when we find that there is only one antecedent

that is shared by all the observed instances of an effect, we infer that it is a cause (book III, ch. VIII, sec. 1; references to Mill hereafter as, e.g., 'III.VIII.1'). Thus we come to believe that hangovers are caused by heavy drinking. According to the Method of Difference, when we find that there is only one prior difference between a situation where the effect occurs and an otherwise similar situation where it does not, we infer that the antecedent that is only present in the case of the effect is a cause (III.VIII.2). If we add sodium to a blue flame, and the flame turns yellow, we infer that the presence of sodium is a cause of the new color, since that is the only difference between the flame before and after the sodium was added. If we once successfully follow a recipe for baking bread, but fail another time because we have left out the yeast and the bread does not rise, we would infer that the yeast is a cause of the rising in the first case. Both methods work by a combination of retention and variation. When we apply the Method of Agreement, we hold the effect constant and try to vary the background as much as we can, and see what stays the same; when we apply the Method of Difference, we vary the effect, and try to hold as much of the background constant as we can, and see what changes.

Mill's methods have a number of attractive features. Many of our inferences are causal inferences, and Mill's methods give a natural account of these. In science, for example, the controlled experiment is a particularly common and self-conscious application of the Method of Difference. The Millian structure of causal inference is often particularly clear in cases of inferential dispute. When you dispute my claim that C is the cause of E, you will often make your case by pointing out that the conditions for Mill's methods are not met, that is, by pointing out C is not the only antecedent common to all cases of E, or that the presence of C is not the only salient difference between a case where E occurs and a similar case where it does not. Mill's methods may also avoid some of the over-permissiveness of other accounts, because of the strong constraints that the requirements of varied or shared backgrounds place on their application. These requirements suggest how our background beliefs influence our inferences, something a good account of inference must do. . . .

Of course, Mill's methods have their share of liabilities, of which I will mention just two. First, they do not themselves apply to unobservable causes or to any causal inferences where the cause's existence, and not just its causal status, is inferred. Second, if the methods are to apply at all, the requirement that there be only a single agreement or difference in antecedents must be seen as an idealization, since this condition is never met in real life. We need principles for selecting from among multiple agreements or similarities those that are likely to be causes, but these are principles Mill does not himself supply. . . . However, Mill's methods can be modified and expanded in a way that may avoid these and other liabilities it faces in its simple form.

This chapter has set part of the stage for an investigation of our inductive practices. I have suggested that many of the problems those practices raise can be set out in a natural way in terms of the underdetermination that is characteristic of inductive inference. The underdetermination of our inferences by our evidence provides the skeptic with his lever, and so poses the problem of justification. It also elucidates the structure of the descriptive problem, and the black box inferences it will take to solve it. I have canvassed several solutions to the problem of description, partly to give a sense of some of our options and partly to suggest just how difficult the problem is. . . .

■ | References

- Chomsky, N. (1965) *Aspects of the Theory of Syntax*, Cambridge, Mass: MIT Press.
- Glimour, C. (1980) *Theory and Evidence*, Princeton: Princeton University Press.
- Goodman, N. (1983) *Fact, Fiction, and Forecast*, 4th edn., Indianapolis: Bobbs-Merrill.
- Hempel, C. (1965) *Aspects of Scientific Explanation*, New York: Free Press.
- Hempel, C. (1966) *Philosophy of Natural Science*, Englewood Cliffs: Prentice-Hall.
- Horwich, D. (1977) *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, L. A. Selby-Bigge and P. H. Niddich (eds.), Oxford: Oxford University Press (1975).
- Kuhn, T. (1970) *The Structure of Scientific Revolutions*, 2nd edn., Chicago: University of Chicago Press.
- Kuhn, T. (1977) *The Essential Tension*, Chicago: University of Chicago Press.
- Mill, J. S. (1904) *A System of Logic*, 8th edn., London: Longmans, Green and Co.
- Skyrms, B. (1986) *Chance and Choice*, 3rd edn., Belmont: Wadsworth.
- Wilbans, B. (1978) *Descartes*, Harmondsworth: Penguin.

The Problem of Induction

1 | The Problem of Induction

According to a widely accepted view . . . the empirical sciences can be characterized by the fact that they use '*inductive methods*', as they are called. According to this view, the logic of scientific discovery would be identical with inductive logic, *i.e.* with the logical analysis of these inductive methods.

It is usual to call an inference 'inductive' if it passes from *singular statements* (sometimes also called 'particular' statements), such as accounts of the results of observations or experiments, to *universal statements*, such as hypotheses or theories.

Now it is far from obvious, from a logical point of view, that we are justified in inferring universal statements from singular ones, no matter how numerous; for any conclusion drawn in this way may always turn out to be false: no matter how many instances of white swans we may have observed, this does not justify the conclusion that *all* swans are white.

The question whether inductive inferences are justified, or under what conditions, is known as *the problem of induction*.

The problem of induction may also be formulated as the question of how to establish the truth of universal statements which are based on experience, such as the hypotheses and theoretical systems of the empirical sciences. For many people believe that the truth of these universal statements is '*known by experience*'; yet it is clear that an account of an experience—of an observation or the result of an experiment—can in the first place be only a singular statement and not a universal one. Accordingly, people who say of a universal statement that we know its truth from experience usually mean that the truth of this universal statement can

somehow be reduced to the truth of singular ones, and that these singular ones are known by experience to be true, which amounts to saying that the universal statement is based on inductive inference. Thus to ask whether there are natural laws known to be true appears to be only another way of asking whether inductive inferences are logically justified.

Yet if we want to find a way of justifying inductive inferences, we must first of all try to establish a *principle of induction*. A principle of induction would be a statement with the help of which we could put inductive inferences into a logically acceptable form. In the eyes of the upholders of inductive logic, a principle of induction is of supreme importance for scientific method: '... this principle', says Reichenbach,^{*} 'determines the truth of scientific theories. To eliminate it from science would mean nothing less than to deprive science of the power to decide the truth or falsity of its theories. Without it, clearly, science would no longer have the right to distinguish its theories from the fanciful and arbitrary creations of the poet's mind.'

Now this principle of induction cannot be a purely logical truth like a tautology or an analytic statement. Indeed, if there were such a thing as a purely logical principle of induction, there would be no problem of induction; for in this case, all inductive inferences would have to be regarded as purely logical or tautological transformations, just like inferences in deductive logic. Thus the principle of induction must be a synthetic statement, that is, a statement whose negation is not self-contradictory but logically possible. So the question arises why such a principle should be accepted at all, and how we can justify its acceptance on rational grounds.

Some who believe in inductive logic are anxious to point out, with Reichenbach, that 'the principle of induction is unreservedly accepted by the whole of science and that no man can seriously doubt this principle in everyday life either'.² Yet even supposing this were the case—for after all, 'the whole of science' might err—I should still contend that a principle of induction is superfluous, and that it must lead to logical inconsistencies.

That inconsistencies may easily arise in connection with the principle of induction should have been clear from the work of Hume; also, that they can be avoided, if at all, only with difficulty. For the principle of induction must be a universal statement in its turn. Thus if we try to regard its truth as known from experience, then the very same problems which occasioned its introduction will arise all over again. To justify it, we should

* Hans Reichenbach (1891–1953) was a leading figure (along with Carl Hempel) in the Berlin school of philosophy of science in the 1920s and early 1930s. Although closely associated with the Vienna Circle, Reichenbach rejected the verifiability principle of meaning and adopted the phrase *logical empiricism* to distinguish his views from those of the logical positivists. His works on space and time, quantum mechanics, probability, and induction have been an important influence on twentieth-century philosophy of science.

have to employ inductive inferences; and to justify these we should have to assume an inductive principle of a higher order; and so on. Thus the attempt to base the principle of induction on experience breaks down, since it must lead to an infinite regress.

Kant tried to force his way out of this difficulty by taking the principle of induction (which he formulated as the 'principle of universal causation') to be '*a priori* valid'. But I do not think that his ingenious attempt to provide an *a priori* justification for synthetic statements was successful.

My own view is that the various difficulties of inductive logic here sketched are insurmountable. So also, I fear, are those inherent in the doctrine, so widely current today, that inductive inference, although not 'strictly valid', can attain some degree of 'reliability' or of 'probability'. According to this doctrine, inductive inferences are 'probable inferences'.¹ 'We have described', says Reichenbach, 'the principle of induction as the means whereby science decides upon truth. To be more exact, we should say that it serves to decide upon probability. For it is not given to science to reach either truth or falsity . . . but scientific statements can only attain continuous degrees of probability whose unattainable upper and lower limits are truth and falsity'.²

At this stage I can disregard the fact that the believers in inductive logic entertain an idea of probability. . . . I can do so because the difficulties mentioned are not even touched by an appeal to probability. For if a certain degree of probability is to be assigned to statements based on inductive inference, then this will have to be justified by invoking a new principle of induction, appropriately modified. And this new principle in its turn will have to be justified, and so on. Nothing is gained, moreover, if the principle of induction, in its turn, is taken not as 'true' but only as 'probable'. In short, like every other form of inductive logic, the logic of probable inference, or 'probability logic', leads either to an infinite regress, or to the doctrine of *apriorism*.

The theory to be developed in the following pages stands directly opposed to all attempts to operate with the ideas of inductive logic. It might be described as the theory of the *deductive method of testing*, or as the view that a hypothesis can only be empirically *tested*—and only *after* it has been advanced.

Before I can elaborate this view (which might be called 'deductivism', in contrast to 'inductivism') I must first make clear the distinction between the *psychology of knowledge* which deals with empirical facts, and the *logic of knowledge* which is concerned only with logical relations. For the belief in inductive logic is largely due to a confusion of psychological problems with epistemological ones. It may be worth noticing, by the way, that this confusion spells trouble not only for the logic of knowledge but for its psychology as well.

2 | Elimination of Psychologism

I said above that the work of the scientist consists in putting forward and testing theories

The initial stage, the act of conceiving or inventing a theory, seems to me neither to call for logical analysis nor to be susceptible of it. The question how it happens that a new idea occurs to a man—whether it is a musical theme, a dramatic conflict, or a scientific theory—may be of great interest to empirical psychology; but it is irrelevant to the logical analysis of scientific knowledge. This latter is concerned not with *questions of fact* (Kant's *quid facti?*), but only with questions of *justification or validity* (Kant's *quid juris?*). Its questions are of the following kind. Can a statement be justified? And if so, how? Is it testable? Is it logically dependent on certain other statements? Or does it perhaps contradict them? In order that a statement may be logically examined in this way, it must already have been presented to us. Someone must have formulated it, and submitted it to logical examination.

Accordingly I shall distinguish sharply between the process of conceiving a new idea, and the methods and results of examining it logically. As to the task of the logic of knowledge—in contradistinction to the psychology of knowledge—I shall proceed on the assumption that it consists solely in investigating the methods employed in those systematic tests to which every new idea must be subjected if it is to be seriously entertained.

Some might object that it would be more to the purpose to regard it as the business of epistemology to produce what has been called a '*rational reconstruction*' of the steps that have led the scientist to a discovery—to the finding of some new truth. But the question is, what, precisely, do we want to reconstruct? If it is the processes involved in the stimulation and release of an inspiration which are to be reconstructed, then I should refuse to take it as the task of the logic of knowledge. Such processes are the concern of empirical psychology but hardly of logic. It is another matter if we want to reconstruct rationally the *subsequent tests* whereby the inspiration may be discovered to be a discovery, or become known to be knowledge. In so far as the scientist critically judges, alters, or rejects his own inspiration we may, if we like, regard the methodological analysis undertaken here as a kind of '*rational reconstruction*' of the corresponding thought-processes. But this reconstruction would not describe these processes as they actually happen: it can give only a logical skeleton of the procedure of testing. Still, this is perhaps all that is meant by those who speak of a '*rational reconstruction*' of the ways in which we gain knowledge

... My view of the matter, for what it is worth, is that there is no such thing as a logical method of having new ideas, or a logical reconstruction of this process. My view may be expressed by saying that every

discovery contains 'an irrational element', or 'a creative intuition', in Bergson's sense.* In a similar way Einstein speaks of the 'search for those highly universal laws . . . from which a picture of the world can be obtained by pure deduction. There is no logical path', he says, 'leading to these . . . laws. They can only be reached by intuition, based upon something like an intellectual love (*Einführung*) of the objects of experience'.⁶

3 | Deductive Testing of Theories

According to the view that will be put forward here, the method of critically testing theories, and selecting them according to the results of tests, always proceeds on the following lines. From a new idea, put up tentatively, and not yet justified in any way—an anticipation, a hypothesis, a theoretical system, or what you will—conclusions are drawn by means of logical deduction. These conclusions are then compared with one another and with other relevant statements, so as to find what logical relations (such as equivalence, derivability, compatibility, or incompatibility) exist between them.

We may if we like distinguish four different lines along which the testing of a theory could be carried out. First there is the logical comparison of the conclusions among themselves, by which the internal consistency of the system is tested. Secondly, there is the investigation of the logical form of the theory, with the object of determining whether it has the character of an empirical or scientific theory, or whether it is, for example, tautological. Thirdly, there is the comparison with other theories, chiefly with the aim of determining whether the theory would constitute a scientific advance should it survive our various tests. And finally, there is the testing of the theory by way of empirical applications of the conclusions which can be derived from it.

The purpose of this last kind of test is to find out how far the new consequences of the theory—whatever may be new in what it asserts—stand up to the demands of practice, whether raised by purely scientific experiments, or by practical technological applications. Here too the procedure of testing turns out to be deductive. With the help of other statements, previously accepted, certain singular statements—which we may call 'predictions'—are deduced from the theory; especially predictions that

* The French philosopher Henri Bergson (1859–1941) attacked materialism and mechanism, rejected science as a complete account of reality, and advocated vitalism, a worldview based on creative forces and intuition. His writings on memory and our subjective experience of time were quite influential, especially on authors such as Marcel Proust, but his postulation of the *élan vital*, a spiritual force that drives biological evolution, discredited Bergson's vitalism in the eyes of most scientists and philosophers.

are easily testable or applicable. From among these statements, those are selected which are not derivable from the current theory, and more especially those which the current theory contradicts. Next we seek a decision as regards these (and other) derived statements by comparing them with the results of practical applications and experiments. If this decision is positive, that is, if the singular conclusions turn out to be acceptable, or *verified*, then the theory has, for the time being, passed its test: we have found no reason to discard it. But if the decision is negative, or in other words, if the conclusions have been *falsified*, then their falsification also falsifies the theory from which they were logically deduced.

It should be noticed that a positive decision can only temporarily support the theory, for subsequent negative decisions may always overthrow it. So long as a theory withstands detailed and severe tests and is not superseded by another theory in the course of scientific progress, we may say that it has 'proved its mettle' or that it is '*corroborated*'.

Nothing resembling inductive logic appears in the procedure here outlined. I never assume that we can argue from the truth of singular statements to the truth of theories. I never assume that by force of 'verified' conclusions, theories can be established as 'true', or even as merely 'probable' . . .

■ | Notes

1 H. Reichenbach, *Erkenntnis* 1, 1930, p. 186 (cf. also p. 64 f.).

2 Reichenbach *ibid.*, p. 67.

3 Cf. J. M. Keynes, *A Treatise on Probability* (1921); O. Külpe, *Vorlesungen über Logik* (ed. by Selz, 1923); Reichenbach (who uses the term 'probability implications'), *Axiomatik der Wahrscheinlichkeitrechnung*, *Mathem. Zeitschr.* 34 (1932); and in many other places.

4 Reichenbach, *Erkenntnis* 1, 1930, p. 186.

5. Liebig (in *Induktion und Deduktion*, 1865) was probably the first to reject the inductive method from the standpoint of natural science; his attack is directed against Bacon. Duhem (in *La Théorie physique, son objet et sa structure*, 1906; English translation by P. P. Wiener: *The Aim and Structure of Physical Theory*, Princeton, 1954) held pronounced deductivist views. (But there are also inductivist views to be found in Duhem's book, for example in the third chapter, Part One, where we are told that only experiment, induction, and generalization have produced Descartes's law of refraction; cf. the English translation, p. 34.) See also V. Kraft, *Die Grundformen der Wissenschaftlichen Methoden*, 1925; and Carnap, *Erkenntnis* 2, 1932, p. 440.

6 Address on Max Planck's 60th birthday. The passage quoted begins with the words, 'The supreme task of the physicist is to search for those highly universal

laws . . . ,’ etc. (quoted from A. Einstein, *Mein Weltbild*, 1934, p. 168; English translation by A. Harris: *The World As I See It*, 1935, p. 125). Similar ideas are found earlier in Liebig, *op. cit.*; cf. also Mach, *Principien der Wärmelehre* (1896), p. 443 ff. The German word ‘*Einfühlung*’ is difficult to translate. Harris translates: ‘sympathetic understanding of experience’.

Rational Prediction

A colleague, to whom I shall refer (quite accurately) as "the friendly physicist," recently recounted the following incident. While awaiting takeoff on an airplane, he noticed a young boy sitting across the aisle holding onto a string to which was attached a helium-filled balloon. He endeavored to pique the child's curiosity. "If you keep holding the string just as you are now," he asked, "what do you think the balloon will do when the airplane accelerates before takeoff?" The question obviously had not crossed the youngster's mind before that moment, but after giving it a little thought, he expressed the opinion that the balloon would move toward the back of the cabin. "I don't think so," said the friendly physicist, "I think it will move forward." The child was now eager to see what would happen when the plane began to move. Several adults in the vicinity were, however, skeptical about the physicist's prediction; in fact, a stewardess offered to wager a miniature bottle of Scotch that he was mistaken. The friendly physicist was not unwilling and the bet was made. In due course, the airplane began to accelerate, and the balloon moved toward the front of the cabin. The child's curiosity was satisfied¹; the theory—that all objects which are free to move will move toward the back of the cabin when the plane accelerates—was falsified; and the friendly physicist enjoyed a free drink.

I have related this anecdote to point out that there are at least three—probably more—legitimate reasons for making predictions. First, we are sometimes curious about future happenings, and we want to satisfy that curiosity without waiting for the events in question to transpire. To do so.

FROM A. Grünbaum and W. C. Salmon, eds., *The Limitations of Deductivism* (Berkeley, Calif.: University of California Press, 1988), 47–60. This article was originally published in the *British Journal for the Philosophy of Science* 32 (1981): 115–125 and incorporates some minor revisions made by the author when the paper was reprinted.

we may make wild guesses, we may employ superstitious methods of prediction, we may appeal to common sense, or we may use more sophisticated scientific theories. Second, we sometimes make predictions for the sake of testing a theory. In the example at hand, the prediction regarding the motion of the balloon was a rather good test of the hypothesis that all objects free to move in the cabin will tend to move toward the rear when the airplane accelerates. The fact that objects heavier than air tend to fall toward the earth when they are unsupported, while objects lighter than air (such as helium-filled balloons) tend to move in the opposite direction, suggests that the behaviour of a helium-filled balloon has a reasonable chance of falsifying the hypothesis about the behavior of all material objects in the air-filled cabin of the accelerating airplane, if it is indeed false. Third, we sometimes find ourselves in situations in which some practical action is required, and the choice of an optimal decision depends upon predicting future occurrences. Although wagering is by no means the only such type of practical decision-making, it is a clear and comprehensible example. We all agree, I take it, that scientific theories often provide sound bases for practical prediction.

A central feature of Sir Karl Popper's philosophy is his thesis concerning the status of induction. Indeed, he begins his book *Objective Knowledge* with the statement: "I think that I have solved a major philosophical problem: the problem of induction. . . . This solution has been extremely fruitful, and it has enabled me to solve a good number of other philosophical problems" (1972, p. 1). His solution, as is well known, involves a complete rejection of induction. This claim has been advanced in many of his writings spanning several decades, and it is reiterated in his autobiography (1974a) and in his "Replies to My Critics" (1974b).

For some time it has seemed to me that the crucial test of an anti-inductivist philosophy of science would be its capacity to deal with the predictive aspects of scientific knowledge. In a paper (Salmon 1968a) presented at the 1965 International Colloquium on Philosophy of Science at Bedford College, London, I attempted to offer a severe challenge to Popper's views concerning induction by posing what I took to be a serious dilemma: On Popper's account, either science embodies essential inductive aspects or else science is lacking in predictive content.² In the published proceedings of the Bedford College Colloquium (Lakatos 1968), J. W. N. Watkins contributed an answer to my critique. He denied that scientific reasoning is inductively infected, and he argued that it can, nevertheless, provide a basis for rational prediction. In Popper's replies to his critics (1974b, pp. 1028–1030), he acknowledges that I have understood his views "fairly well," and he endorses Watkins's response. I take this as evidence that we have located a genuine disagreement—one which is reasonably free from purely verbal disputes or out-and-out misrepresentations—regarding Popper's anti-inductivist stand. The question involves what Popper calls "the pragmatic problem of induction." It is this

issue that I want to pursue in the present paper; it concerns the problem of rational prediction. Although the issue may appear to be rather narrow, it seems to me to have pivotal importance with regard to the assessment of Popper's deductivism.

Let me attempt to formulate the basic difficulty as I see it. In its very simplest terms, Popper's account of scientific knowledge involves generalizations and their observational tests. If we find a *bona fide* counterexample to a generalization, we can say that it has been deductively refuted. To be sure, as Popper explicitly acknowledges, there may be difficulties in some cases in determining whether certain observations constitute genuine counterexamples to a generalization, but that does not undermine the claim that a genuine counterexample yields a deductive refutation. According to Popper, negative instances provide rational grounds for rejecting generalizations. If, however, we make observations and perform tests, but no negative instance is found, all we can say deductively is that the generalization in question has not been refuted. In particular, positive instances do not provide confirmation or inductive support for any such unrefuted generalization. At this stage, I claim, we have no basis for rational prediction. Taken in themselves, our observation reports refer to past events, and consequently they have no predictive content. They say nothing about future events. If, however, we take a general statement as a premise, and conjoin to it some appropriate observation statements about past or present events, we may be able to deduce a conclusion which says something about future occurrences and that, thereby, has predictive content. Popper himself gives this account of *the logic of prediction* (1947b, p. 1030).

The problem of rational prediction concerns the status of the general premise in such an argument. One may claim, as Popper does, that we ought not to use a generalization that has actually been refuted as a premise in a predictive argument of this sort, for we are justified in regarding it as false. We ought not to employ premises which are known to be false if we hope to deduce true predictions. The exclusion of refuted generalizations does not, however, tell us what general premise should be employed. Typically there will be an infinite array of generalizations which are compatible with the available observational evidence, and that are therefore, as yet, unrefuted. If we were free to choose arbitrarily from among all the unrefuted alternatives, we could predict anything whatever. If there were no rational basis for choosing from among all of the unrefuted alternatives, then, as I think Popper would agree, there would be no such thing as rational prediction. We are not in this unfortunate situation, Popper contends, for we do have grounds for preferring one unrefuted generalization to another: "My solution of the logical problem of induction was that we may have *preferences* for certain of the competing conjectures; that is, for those which are highly informative and which so far have stood up to eliminative criticism" (1974b, p. 1024). Popper's concept of corrob-

oration is designed to measure the manner in which conjectures have stood up to severe criticism, including severe testing. This, I take it, is the crucial thesis—that *there is a rational basis for preferring one unrefuted generalization to another for use in a predictive argument*. If that is correct, then Popper can legitimately claim to have solved the problem of rational prediction.

If we are going to talk about preference among generalizations, then we have to be quite explicit about the purpose for which the generalization is to be used. In this context, we are discussing prediction, so the preference must be in relation to predictive capability. As Popper rightly insists, any generalization we choose will have predictive import in the sense that it will make statements about future events—more precisely, in a predictive argument as characterized above, it yields conclusions about future occurrences. But since all of the various unrefuted generalizations have predictive content in that sense, we must still ask on what basis the predictive content of one conjecture is rationally preferable to that of another conjecture.

At this stage of the discussion, it is important to recall the point of the opening story, namely, that predictions are made for various purposes. Thus, even if we agree that we want to select a generalization for predictive purposes, we must still specify what type of prediction is involved. Popper explicitly acknowledges (1974*b*, pp. 1024–1025) that there are two types of preference, “the theoretician’s preference” and that of “the man of practical action.” As I understand Popper’s view, the theoretician is interested in formulating bold conjectures which have high content and in subjecting them to severe tests. Insofar as the theoretician is mainly interested in explanations of known phenomena, he may not be much involved in making any sorts of predictions. I suppose we might distinguish the theoretician’s explanatory preference from the theoretician’s predictive preference, recognizing that there is bound to be a close connection between preferences of these two kinds. When the theoretician is actually involved *qua theoretician* in making predictions, the purpose is to devise (and, perhaps, to instruct the experimentalist on how to conduct) a severe test. The purpose of predictions made in this theoretical context is to gain information that is useful in the evaluation of scientific theories. If the chief value of the scientific theories is explanatory, then it is not at all clear that a primary desideratum of the predictive argument is to arrive at a true prediction. As Popper has emphasized, and as all of us know, a false prediction can be valuable, since the realization (on the basis of observation) that it is false can be highly informative.

Having briefly characterized theoretical preference, let us now focus attention upon the kind of preference which is pertinent to the practical context, with special attention to the kinds of predictions which play a role in practical decision making. As I have remarked above, Popper claims that for theoretical purposes we prefer theories which are highly

corroborated to those that are less well corroborated. I do not think this claim is unproblematic, but I do not propose arguing the matter here. My aim is to emphasize that, even if we are entirely justified in letting such considerations determine our theoretical preferences, it is by no means obvious that we are justified in using them as the basis for our preferences among generalizations which are to be used for prediction in the practical decision-making context. Popper and Watkins have maintained, however, that corroboration should play a crucial role in determining both theoretical preference and practical preference.

Since scientific theories are used for both theoretical and practical purposes—including prediction—and since, according to Popper, theory preference is based upon corroboration, I had mistakenly inferred (prior to 1968) that the appraisal of a theory in terms of corroboration must imply some attempt at an appraisal of the theory with respect to its future performance. If that were Popper's thesis, I had argued, then corroboration must involve some element of induction (or nondemonstrative inference of some sort), for past performance of the theory is taken to constitute a basis for some sort of claim about future performance. However, I have since been informed by Watkins (1968) and Popper (1974a) that I had misconstrued Popper's view. Statements about the corroboration of theories are no more than appraisals of their past performances, corroboration statements hold no predictions with respect to future performance. If they did, they would be inductive (as I had claimed); but they are not inductive, so they cannot be predictive.

This view of corroboration holds serious difficulties. Watkins and Popper agree, I take it, that statements that report observations of past and present events do not, in and of themselves, have any predictive content. Moreover, they maintain, statements about the corroboration of conjectures do not, in and of themselves, have any predictive content. Conjectures, hypotheses, theories, generalizations—call them what you will—do have predictive content. The problem is that there are many such statements, rich in predictive content, which make incompatible predictive claims when conjoined with true statements about past and present occurrences. The fact that a general statement has predictive content does not mean that what it says is true. In order to make a prediction, one must choose a conjecture that has predictive content to serve as a premise in a predictive argument. In order to make a *rational* prediction, it seems to me, one must make a *rational* choice of a premise for such an argument. But from our observational evidence and from the statements about the corroboration of a given conjecture, no predictive appraisal follows. Given two conjectures which, in a particular situation, will lead to incompatible predictions, and given the corroboration ratings of these two hypotheses, *nothing follows* about their comparative predictive capacities. Thus, it seems to me, corroboration—the ground for theoretical preference—furnishes no rational basis for preference of one conjecture to another for

purposes of practical prediction. I am not complaining that we are not told for sure that one will make a correct prediction and that the other will not. I am complaining that no rational basis whatever has been furnished for a preference of this type.

In his reply to my Bedford College paper, Watkins acknowledges that there is an important distinction between theoretical and practical preferences, and he further acknowledges that the two kinds of appraisal may have quite different bases:

Now our methods of hypothesis-selection in practical life should be well suited to our practical aims, just as our methods of hypothesis-selection in theoretical science should be well suited to our theoretical aims; and the two kinds of method may very well yield different answers in a particular case (1968, p. 65).

He goes on to explain quite correctly how utility considerations may bear upon the practical situation. Then he considers the case in which utility does not play a decisive role:

Now suppose that, for a particular agent, the mutually incompatible hypotheses h_1 and h_2 are on a par utility-wise, and that in the situation in which he finds himself, he has got to act since 'inaction' would itself be one mode of action. Then if h_1 is the only alternative to h_2 before him, he has to choose one of them. Then it would be rational for him to choose the better corroborated one, the one which has withstood the more severe criticism, since he has nothing else to go on. (Pp. 65–66).

Watkins offers no further argument for supposing that corroboration provides a rational basis for *practical* preference. Moreover, the hint of an argument which he does supply appeals to a false premise. The agent does have other things "to go on." He could decide between the two hypotheses by the flip of a coin. He could count the numbers of characters in each of the two hypotheses in the particular formulation given, and choose the one that has fewer. He could choose the hypothesis which comes first lexicographically in the given formulation. What Watkins is suggesting, it seems to me, is not that the agent has "nothing else to go on" but rather that he has no other *rational* basis for preference. But such an argument would be patently question begging. Even if all other bases for choice were irrational, it would not follow that the one cited by Watkins is *ipso facto* rational. Indeed, if we take seriously Popper's statement, "I regarded (and I still regard) the degree of corroboration of a theory merely as a critical report on the quality of past performance: it could not be used to predict future performance" (1974a, p. 82), it is hard to see how corroboration can supply a rational basis for preference of a theory for *purposes of practical prediction*.

Whether my criticism of Popper's position is correct or incorrect, the issue I am raising has fundamental importance. For if it should turn out that Popper could not provide a tenable account of rational prediction, then—given his persistent emphasis upon objectivity and rationality—we could hardly credit his claim to have solved the problem of induction. Moreover, in his replies to his critics, Popper acknowledges the issue. With the comment, "Our corroboration statements have no predictive import, although they motivate and justify our *preference* for some theory over another" (1974b, pp. 1029–1030), he endorses the answer Watkins had furnished. Since I am not attempting to deal with the psychological problem of induction, I shall not dispute the claim that corroboration may *motivate* the preference of one theory to another. What I want to see is how corroboration could *justify* such a preference. Unless we can find a satisfactory answer to that question, it appears to me that we have no viable theory of *rational* prediction, and no adequate solution to the problem of induction.

In *Objective Knowledge*, Popper offers an answer to the basic question which seems closely related to that of Watkins:

[A] *pragmatic belief in the results of science* is not irrational, because there is nothing more 'rational' than the method of critical discussion, which is the method of science. And although it would be irrational to accept any of its results as certain, there is nothing 'better' when it comes to practical action: there is no alternative method which might be said to be more rational. (1972, p. 27).

This response appears to miss the point. The question is not whether other methods—for example, astrology or numerology—provide more rational approaches to prediction than does the scientific method. The question is whether the scientific approach provides a more rational basis for prediction, for purposes of practical action, than do these other methods. The position of the Humean skeptic would be, I should think, that none of these methods can be shown either more or less rational than any of the others. But if every method is equally lacking in rational justification, then there is no method which can be said to furnish a rational basis for prediction, for any prediction will be just as unfounded rationally as any other. If the Humean skeptic were right, we could offer the following parallel claim. A pragmatic belief in the predictions found in Chinese fortune cookies is not irrational, for there is nothing more rational. . . .

In his replies to his critics, Popper again addressed the problem, and he came more firmly to grips with it:

But every action presupposes a set of expectations, that is, of theories about the world. Which theory shall the man of action choose? Is there such a thing as a *rational choice*?

This leads us to the *pragmatic problems of induction*, which to start with, we might formulate thus:

- (a) Upon which theory should we rely for practical action, from a rational point of view?
- (b) Which theory should we prefer for practical action, from a rational point of view?

My answer to (a) is: from a rational point of view, we should not 'rely' on any theory, for no theory has been shown to be true, or can be shown to be true (or 'reliable').

My answer to (b) is: we should *prefer* the best tested theory as a basis for action.

In other words, there is no 'absolute reliance'; but since we *have* to choose, it will be 'rational' to choose the best tested theory. This will be 'rational' in the most obvious sense of the word known to me: the best tested theory is the one which, in the light of our *critical discussion*, appears to be the best so far; and I do not know of anything more 'rational' than a well-conducted critical discussion (1974b, p. 1025)

Let us not be seduced by honeyed words. If we wish to claim that a theory "appears to be the best so far," we must ask, "Best for what purpose—*theoretical explanation* or *practical prediction*?" Since it is "the best tested theory" and it has been subjected to "critical discussion," then, in the light of the many statements by Popper and others about the lack of predictive import of corroboration, we must conclude, I believe, that the answer is, "Best for *theoretical explanation*." Perhaps I am being unduly obtuse, but I cannot see that any reason has been provided for supposing that such a theory is best for *practical prediction*.

I must confess to the feeling that we have been given the runaround. We begin by asking how science can possibly do without induction. We are told that the aim of science is to arrive at the best explanatory theories we can find. When we ask how to tell whether one theory is better than another, we are told that it depends upon their comparative ability to stand up to severe testing and critical discussion. When we ask whether this mode of evaluation does not contain some inductive aspect, we are assured that the evaluation is made wholly in terms of their comparative success up to now; but since this evaluation is made entirely in terms of past performance, it escapes inductive contamination because it lacks predictive import. When we then ask how to select theories for purposes of rational prediction, we are told that we should prefer the theory which is "best tested" and that "in the light of our *critical discussion*, appears to be the best so far," even though we have been explicitly assured that testing and critical discussion have no predictive import. Popper tells us, "I do

not know of anything more 'rational' than a well-conducted critical discussion." I fail to see how it could be rational to judge theories *for purposes of prediction* in terms of a criterion which is emphatically claimed to be lacking in predictive import.³

Fearing that the point of his initial argument may have been missed, Popper attempts another formulation:

Let us forget momentarily about what theories we 'use' or 'choose' or 'base our practical actions on', and consider only the resulting *proposal* or *decision* (to do X; not to do X; to do nothing; or so on). Such a proposal can, we hope, be rationally criticized; and if we are rational agents we will want it to survive, if possible, the most testing criticism we can muster. *But such criticism will freely make use of the best tested scientific theories in our possession.* Consequently any proposal that ignores these theories (where they are relevant, I need hardly add) will collapse under criticism. Should any proposal remain, it will be rational to adopt it.

This seems to me all far from tautological. Indeed, it might well be challenged by challenging the italicized sentence in the last paragraph. Why, it might be asked, does rational criticism make use of the best tested although highly unreliable theories? The answer, however, is exactly the same as before. Deciding to criticize a practical proposal from the standpoint of modern medicine (rather than, say, in phrenological terms) is itself a kind of 'practical' decision (anyway it may have practical consequences). Thus the rational decision is always: adopt critical methods which have themselves withstood severe criticism. (1974b, pp. 1025-1026).

I have quoted Popper *in extenso* to try to be quite sure not to misunderstand his answer. The italicized sentence in the first paragraph raises precisely the question which seems to me crucial. In the second paragraph, Popper admits the legitimacy of the question, and he offers an answer. When he says, "The answer . . . is exactly the same as before. . . . [T]he rational decision is always: adopt critical methods which have themselves withstood severe criticism," he seems to be saying that we should adopt his methodological recommendations, because they have "withstood severe criticism." But his answer is inappropriate in this context because our aim is precisely to subject his philosophical views, in the best Popperian spirit, to severe criticism.

In my reply to Watkins, I said, "Watkins acknowledges . . . that corroboration does have predictive import in practical decision making" (1968b, p. 97). Popper has objected to this way of putting the matter: "[O]ur theories do have predictive import. Our corroboration statements have no predictive import, although they motivate and justify our preference for some theory or other" (1974b, pp. 1029-1030). Let us grant that corroboration statements have no predictive content—indeed, that they are

analytic, as Watkins remarks (1968, p. 63)—and that theories are the kinds of statements that do have predictive content. It does not follow, as Popper has claimed, that corroboration has no predictive import. The distinction between predictive content and predictive import is no mere verbal quibble; a fundamental substantive point is at issue. Statements whose consequences refer to future occurrences may be said to have predictive content; rules, imperatives, and directives are totally lacking in predictive content because they do not entail any statements at all. Nevertheless, an imperative—such as “No smoking, please”—may have considerable predictive import, for it may effectively achieve the goal of preventing the occurrence of smoking in a particular room in the immediate future.

Since corroboration, in some cases at least, provides the basis for deciding which theory (with its predictive content) is to be used for the purpose of making practical predictions, it seems to me that corroboration, even if it is lacking in predictive content, does have enormous predictive import. Perhaps this point can be put more clearly in the following way. Statements assessing the corroboration of theories have no predictive content, as Popper, Watkins, and others maintain. The directive—to choose more highly corroborated theories in preference to theories that are less well corroborated for purposes of practical prediction—has considerable predictive import. The problem, which it seems to me the anti-inductivists have failed to solve, is how to vindicate this directive for making predictions.⁴ Without some sort of vindication for this directive, the problem of rational prediction remains unresolved.

I have wondered why it would seem evident to Popper that corroboration, as he construes it, should provide a guide to rational prediction. In his autobiography, he gives what appear to be indications of an answer.

I regarded (and I still regard) the degree of corroboration of a theory merely as a critical report on the quality of past performance: it *could not be used to predict future performance*. . . . When faced with the *need to act*, on one theory or another, the rational choice was to act on that theory—if there was one—which so far had stood up to criticism better than its competitors had: there is no better idea of rationality than that of a readiness to accept criticism. Accordingly, the degree of corroboration of a theory was a rational guide to practice. (1974a, p. 82)

A further elaboration of the theme informs us that

when we think we have found an approximation to the truth in the form of a scientific theory which has stood up to criticism and to tests better than its competitors, we shall, as realists, accept it as a basis for practical action, simply because we have nothing better (or nearer to the truth). (ibid., pp. 120–121)

Realism is a position to which Popper has adhered since the time of his earliest philosophical activity; near the beginning of his autobiography he tells us that "a realist who believes in an 'external world' necessarily believes in the existence of a cosmos rather than a chaos; that is, in regularities" (ibid., p. 14). Thus, I am led to conjecture, it may be that Popper's adherence to the thesis that corroboration can provide a basis for rational prediction rests ultimately upon his realism, which embodies a version of a principle of uniformity of nature. If this suggestion is correct, we can still legitimately wonder whether Popper's epistemology is as far from traditional inductivism as he would have us believe.

To conclude this discussion, I should like to recall the point of my opening anecdote. It seems to me incorrect to suppose that the only concern of *theoretical* science is to make bold explanatory conjectures that can be tested and criticized. It is a mistake, I believe, to suppose that all prediction, aside from that involved in the testing of theories, is confined to contexts in which practical action is at stake. Theoretical science furnishes both explanations and predictions. Some of these predictions have practical consequences and others do not. When, for example, scientists assembled the first man-made atomic pile under the West Stands at the University of Chicago, they had to make a prediction as to whether the nuclear chain reaction they initiated could be controlled, or whether it would spread to surrounding materials and engulf the entire city—and perhaps the whole earth—in a nuclear holocaust. Their predictions had both theoretical and practical interest. Contemporary cosmologists, for another example, would like to *explain* certain features of our universe in terms of its origin in a "big bang"; many of them are trying to *predict* whether it will end in a "big crunch." In this case, the predictive question seems motivated by pure intellectual curiosity, quite unattached to concerns regarding practical decision making. Whether a helium-filled balloon will move forward in the cabin of an airplane when the airplane accelerates, whether a nuclear chain reaction—once initiated—will run out of control, and whether the universe will eventually return to a state of high density are all matters of legitimate scientific concern.

In this paper, I have attempted to argue that pure deductivism could not do justice to the problem of rational prediction in contexts of practical decision making. If we ask whether Popperian deductivism can adequately account for scientific predictions of the more theoretical varieties, then I suspect that we would have to go through all of the preceding arguments once more. The net result would be, I think, that science is inevitably inductive in matters of intellectual curiosity as well as practical prediction. It may be possible to excise all inductive ingredients from science, but if the operation were successful, the patient (science), deprived of all predictive import, would die.⁵

Notes

1. His curiosity regarding *what* would happen was satisfied, though not his curiosity as to *why*.
2. Similar themes were developed in Salmon 1967, chap. 2, sec. 3.
3. The argument advanced in this paragraph bears a strong resemblance, I think, to one developed in Grünbaum 1976; see esp. p. 246.
4. This felicitous reformulation was suggested by Abner Shimony (if I did not misunderstand him) in the discussion following my presentation at the Popper Symposium.
5. A version of this paper was presented orally at the Symposium on the Philosophy of Sir Karl Popper, London School of Economics, July 14–16, 1980. This material is based upon work supported by the National Science Foundation (U.S.A.) under Grant No. SES-7809146.

References

- Grünbaum, Adolf. 1976. "Is Falsifiability the Touchstone of Scientific Rationality? Karl Popper versus Inductivism." In R. S. Cohen, P. K. Feyerabend, and M. W. Wartofsky, eds., *Essays in Memory of Imre Lakatos*, pp. 213–252. Dordrecht: Reidel.
- Lakatos, Imre, ed. 1968. *The Problem of Inductive Logic*. Amsterdam: North-Holland.
- Popper, Karl R. 1972. *Objective Knowledge*. Oxford: Clarendon Press.
- . 1974a. "Autobiography." In P. A. Schilpp, ed., *The Philosophy of Karl Popper*, pp. 3–181. LaSalle, Ill.: Open Court.
- . 1974b. "Replies to My Critics." In P. A. Schilpp, ed., *The Philosophy of Karl Popper*, pp. 961–1197. LaSalle, Ill.: Open Court.
- Salmon, Wesley C. 1967. *The Foundations of Scientific Inference*. Pittsburgh, Pa.: University of Pittsburgh Press. Originally published in R. C. Colodny, ed., *Mind and Cosmos*, pp. 135–275. Pittsburgh, Pa.: University of Pittsburgh Press, 1966.
- . 1968a. "The Justification of Inductive Rules of Inference." In Lakatos 1968, pp. 24–43.
- . 1968b. "Reply." In Lakatos 1968, pp. 74–97.
- Watkins, J. W. N. 1968. "Non-Inductive Corroboration." In Lakatos 1968, pp. 61–66.

Criteria of Confirmation and Acceptability

... A favorable outcome of even very extensive and exacting tests cannot provide conclusive proof for a hypothesis, but only more or less strong evidential support, or confirmation. How strongly a hypothesis is supported by a given body of evidence depends on various characteristics of the evidence, which we will consider presently. In appraising what might be called the scientific acceptability or credibility of a hypothesis, one of the most important factors to consider is, of course, the extent and the character of the relevant evidence available and the resulting strength of the support it gives to the hypothesis. But several other factors have to be taken into account as well; these, too, will be surveyed in this chapter. We shall at first speak in a somewhat intuitive manner of more or less strong support, of small or large increments in confirmation, of factors that increase or decrease the credibility of a hypothesis, and the like. At the end of the chapter, we will briefly consider whether the concepts here referred to admit of a precise quantitative construal.

1 | Quantity, Variety, and Precision of Supporting Evidence

In the absence of unfavorable evidence, the confirmation of a hypothesis will normally be regarded as increasing with the number of favorable test findings. For example, each new Cepheid variable whose period and luminosity are found to conform to the Leavitt-Shapley law will be consid-

FROM Carl G. Hempel, *Philosophy of Natural Science* (Englewood Cliffs, N.J.: Prentice-Hall, 1966), 33-46.

ered as adding to the evidential support of the law.* But broadly speaking, the increase in confirmation effected by one new favorable instance will generally become smaller as the number of previously established favorable instances grows. If thousands of confirmatory cases are already available, the addition of one more favorable finding will raise the confirmation but little.

This remark must be qualified, however. If the earlier cases have all been obtained by tests of the same kind, but the new finding is the result of a different kind of test, the confirmation of the hypothesis may be significantly enhanced. For the confirmation of a hypothesis depends not only on the quantity of the favorable evidence available, but also on its variety: the greater the variety, the stronger the resulting support.

Suppose, for example, that the hypothesis under consideration is Snell's law, which states that a ray of light traveling obliquely from one optical medium into another is refracted at the separating surface in such a way that the ratio, $\sin \alpha / \sin \beta$, of the sines of the angles of incidence and of refraction is a constant for any pair of media. Compare now three sets of 100 tests each. In the first set, the media and the angle of incidence are kept constant: in each experiment, the ray passes from air into water at an angle of incidence of 30° ; the angle of refraction is measured. Suppose that in all cases, $\sin \alpha / \sin \beta$ does have the same value. In the second set, the media are kept constant, but the angle α is varied: light passes from air into water at varying angles, β is measured. Again, suppose that $\sin \alpha / \sin \beta$ has the same value in all cases. In the third set, both the media and the angle α are varied: 25 different pairs of media are examined: for each pair, four different angles α are used. Suppose that for each pair of media, the four associated values of the ratio $\sin \alpha / \sin \beta$ are equal, while the ratios associated with different pairs have different values.

Each test set then presents a class of favorable outcomes, since the ratios associated with any particular pair of media are found to be equal, as implied by Snell's law. But the third set, which offers the greatest variety of positive instances, will surely be regarded as supporting the law much more strongly than the second, which provides supporting instances of much more limited variety, and the first set, it will be agreed, lends even

* Cepheid variables are stars that fluctuate in brightness with a constant period of one to fifty days. (They are called Cepheids after the prototype Delta Cephei, a variable star discovered in 1784.) The Leavitt-Shapley law is named in honor of the Americans astronomers Henrietta Swan Leavitt and Harlow Shapley, who discovered that the period of Cepheids is proportional to their intrinsic luminosity. Because the apparent brightness of any object falls off as the inverse square of its distance from the observer, the law can be used to calculate the distance to remote galaxies by observing the period and apparent brightness of the Cepheids they contain. To confirm the Leavitt-Shapley law by observing Cepheids requires an independent way of measuring distance to those stars. For stars that are not too far from the earth, this can be done by parallax measurements.

less strong support to the general law. In fact, it might seem that in the first set, the same experiment is performed over and over again, and that the positive outcome in all 100 cases can support the hypothesis no more strongly than do the first two tests in the set, which bear out the constancy of the ratio. But this idea is mistaken. What is repeated here 100 times is not literally the same experiment, for the successive performances differ in many respects, such as the distance of the apparatus from the moon, perhaps the temperature of the light source, the atmospheric pressure, and so on. What is "kept the same" is simply a certain set of conditions, including a fixed angle of incidence and one particular pair of media. And even if the first two or more measurements under these circumstances yield the same value for $\sin \alpha / \sin \beta$, it is logically quite possible that subsequent tests under the specified circumstances should yield different values for the ratio. Thus even here, repeated tests with favorable outcome add to the confirmation of the hypothesis—though much less so than do tests that cover a wider variety of instances.

... Scientific theories are often supported by empirical findings of amazing variety. Newton's theory of gravitation and of motion implies, for example, the laws for free fall, for the simple pendulum, for the motion of the moon about the earth and of the planets about the sun, for the orbits of comets and of man-made satellites, for the motion of double stars about each other, for tidal phenomena, and many more. And all the diverse experimental and observational findings that bear out those laws lend support to Newton's theory.

The reason why diversity of evidence is so important a factor in the confirmation of a hypothesis might be suggested by the following consideration, which refers to our example of various tests for Snell's law. The hypothesis under test—let us call it *S* for short—refers to *all* pairs of optical media and asserts that for any pair, the ratio $\sin \alpha / \sin \beta$ has the same value for *all* associated angles of incidence and of refraction. Now, the more widely a set of experiments ranges over the diverse possibilities here covered, the greater will be the chances of finding an unfavorable instance if *S* should be false. Thus, the first set of experiments may be said to test more specifically a hypothesis *S*₁ that expresses only a small part of Snell's law—namely, that $\sin \alpha / \sin \beta$ has the same value whenever the optical media are air and water and α is 30°. Hence, if *S*₁ should be true, but *S* false, the first kind of test will never disclose this. Similarly, the second set of experiments tests a hypothesis *S*₂, which asserts distinctly more than *S*₁ but still not nearly as much as *S*—namely, that $\sin \alpha / \sin \beta$ has the same value for all angles α and the associated angles β if the media involved are air and water. Hence, if *S*₂ should be true, but *S* false, a test set of the second kind would never disclose this. Thus, the third set of experiments might be said to test Snell's law more thoroughly than the other two; an entirely favorable outcome accordingly lends stronger support to it.

As an additional illustration of the power of diversified evidence, we might note that if the diversity of the evidence is still further increased by varying the temperature of the optical media or by using monochromatic light of different wavelengths, then Snell's law in the classical form cited above is in fact found to be false.

But have we not overstated the case for diversified evidence? After all, some ways of increasing variety would be regarded as pointless, as incapable of raising the confirmation of a hypothesis. This verdict would apply, for example, if in our first test set for Snell's law the variety were increased by having the experiment performed at different places, during different phases of the moon, or by experimenters with different eye color. But to try such variations would not be unreasonable if as yet we had no knowledge, or only extremely limited knowledge, of what factors are likely to affect optical phenomena. At the time of the Puy-de-Dôme experiment, for example, the experimenters had no very definite ideas of what factors other than altitude might affect the length of the mercury column in the barometer, and when Pascal's brother-in-law and his associates performed the Torricelli experiment on the mountaintop and found the mercury column over three inches shorter than it had been at the foot of the mountain, they decided to repeat the experiment then and there, changing the circumstances in various ways.* As Périer says in his report: "I therefore tried the same thing five times more, with great accuracy, at different places on the top of the mountain, once under cover in the little chapel which is there, once exposed, once in a shelter, once in the wind, once in good weather, and once during the rain and the mists which came over us sometimes, having taken care to get rid of the air in the tube every

* The Puy-de-Dôme experiment was performed at the instigation of Blaise Pascal (1623–62) in 1648. The first mercury barometer had been constructed by Galileo's pupil, Torricelli, in 1644. Torricelli, like Galileo, believed that the barometer acts as a simple mechanical balance, the twenty-nine-inch column of mercury and a vacuum on one side being supported by the pressure of the atmosphere on the other. Aristotelians insisted that the entire effect is nonmechanical and due to nature's abhorrence of a vacuum (*horror vacui*). Prior to the invention of the air pump (in the 1650s), it was hard to rule out the Aristotelian theory, since a conclusive refutation would require placing the entire apparatus in a sealed chamber, reducing the pressure (by pumping out the air), and then observing a steady decrease in the height of the mercury column. But Pascal realized that, even without an air pump, a crucial test of the rival theories was possible because atmospheric pressure decreases as one ascends a mountain. Thus, he persuaded his brother-in-law Périer, who lived near the Puy-de-Dôme mountain in central France, to perform the famous experiment alluded to by Hempel. For an eyewitness report of the experiment, see W. F. Magie, *A Source Book in Physics* (Cambridge, Mass.: Harvard University Press, 1963), 70–75. There is a good account of this and Pascal's other brilliant experiment—the "vacuum in a vacuum"—in Richard S. Westfall, *The Construction of Modern Science: Mechanisms and Mechanics* (New York: John Wiley and Sons, 1971), 43–49.

time; and in all these trials there was found the same height of the quicksilver . . . ; this result fully satisfied us."¹

Thus, the qualification of certain ways of varying the evidence as important and of other ways as pointless is based on the background assumptions we entertain—perhaps as a result of previous research—concerning the probable influence of the factors to be varied upon the phenomenon with which the hypothesis is concerned.

And sometimes when such background assumptions are questioned and experimental variations are accordingly introduced which, on the generally accepted view, are pointless, a revolutionary discovery may be the outcome. This is illustrated by the recent overthrow of one of the basic background assumptions of physics, the principle of parity. According to this principle, the laws of nature are impartial between right and left; if a certain kind of physical process is possible (i.e., if its occurrence is not precluded by the laws of nature), then so is its mirror image (the process as seen in a reflecting mirror), where right and left are interchanged. In 1956, Yang and Lee, who were trying to account for some puzzling experimental findings concerning elementary particles, suggested that the principle of parity is violated in certain cases, and their bold hypothesis soon received clear experimental confirmation.*

Sometimes, a test can be made more stringent, and its result the more weighty, by increasing the precision of the procedures of observation and measurement it involves. Thus, the hypothesis of the identity of inertial and gravitational mass—supported, for example, by the equality of the accelerations shown in free fall by bodies of different chemical constitution—has recently been re-examined with extremely precise methods; and the results, which have so far borne out the hypothesis, have greatly strengthened its confirmation.[†]

* The first experiment to confirm the Lee-Yang conjecture that parity is not conserved in weak interactions was performed in 1956–57 by a team led by Chien-Shiung Wu, professor of physics at Columbia University. For a clear account of the essential details of her experiment, see Martin Gardner, *The New Ambidextrous Universe*, 3d rev. ed. (New York: W. H. Freeman and Company, 1990), ch. 22.

† Hempel is referring to the experiments begun in 1959 by Robert H. Dicke and his colleagues at Princeton University, who verified to an accuracy of about 1 part in 10^{11} that gravity produces the same acceleration in all bodies regardless of their chemical composition. See Robert H. Dicke, "The Eötvös Experiment," *Scientific American* 205 (1961): 84–94. The hypothesis that inertial mass and gravitational mass are identical is usually referred to as the *principle of equivalence* and is one of the central axioms of Einstein's general theory of relativity.

2 | Confirmation by "New" Test Implications

When a hypothesis is designed to explain certain observed phenomena, it will of course be so constructed that it implies their occurrence; hence, the fact to be explained will then constitute confirmatory evidence for it. But it is highly desirable for a scientific hypothesis to be confirmed also by "new" evidence—by facts that were not known or not taken into account when the hypothesis was formulated. Many hypotheses and theories in natural science have indeed received support from such "new" phenomena, with the result that their confirmation was considerably strengthened.

The point is well illustrated by an example that dates back to the last quarter of the nineteenth century, when physicists were searching for inherent regularities in the profusion of lines that had been found in the emission and absorption spectra of gases. In 1885, a Swiss school teacher, J. J. Balmer, proposed a formula that he thought expressed such a regularity for the wavelengths of a series of lines in the emission spectrum of hydrogen. On the basis of measurements that Ångström had made of four lines in that spectrum, Balmer constructed the following general formula:

$$\lambda = b \frac{n^2}{n^2 - 2^2}$$

Here, b is a constant, whose value Balmer determined empirically as 3645.6 Å, and n is an integer greater than 2. For $n = 3, 4, 5$, and 6, this formula yields values that agree very closely with those measured by Ångström; but Balmer was confident that the other values, too, would represent wavelengths of lines yet to be measured—or even yet to be found—in the hydrogen spectrum. He was unaware that some further lines had already been noted and measured. By now, 35 consecutive lines in the so-called Balmer series for hydrogen have been ascertained, and all of these have wavelengths that agree well with the values predicted by Balmer's formula.²

It is hardly surprising that such striking confirmation by correctly predicted "new" facts greatly enhances the credence we will be prepared to give to a hypothesis. A puzzling question arises in this context. Suppose for a moment that Balmer's formula had been constructed only after all the 35 lines now recorded in the series had been carefully measured. In this fictitious case, then, exactly the same experimental findings would be available that have in fact been obtained by measurements made in part before, and in much larger part after, the constructions of the formula. Should that formula be considered as less well confirmed in the fictitious case than in the actual one? It might seem reasonable to answer in the

affirmative, on these grounds: for *any* given set of quantitative data, it is possible to construct a hypothesis that covers them, just as for any finite set of points, it is possible to draw a smooth curve that contains them all. Thus, there would be nothing very surprising about the construction of Balmer's formula in our fictitious case. What is remarkable, and does lend weight to a hypothesis, is its fitting "new" cases: and Balmer's hypothesis has this accomplishment to its credit in the actual case, but not in the fictitious one. But this argument could be met with the reply that even in the fictitious case, Balmer's formula is not just some otherwise arbitrary hypothesis that is rigged to fit the 35 measured wavelengths: it is, rather, a hypothesis of striking formal simplicity; and the very fact that it subsumes those 35 wavelengths under a mathematically simple formula should lend it much higher credibility than could be accorded to a very complex formula fitting the same data. To state the idea in geometrical terms: if a set of points representing the results of measurements can be connected by a simple curve, we have much greater confidence in having discovered an underlying general law than if the curve is complicated and shows no perceptible regularity. (This notion of simplicity will be further considered, later on in this chapter.) Besides, from a logical point of view, the strength of the support that a hypothesis receives from a given body of data should depend only on what the hypothesis asserts and what the data are: the question of whether the hypothesis or the data were presented first, being a purely historical matter, should not count as affecting the confirmation of the hypothesis. This latter conception is certainly implicit in recently developed statistical theories of testing and also in some contemporary logical analyses of confirmation and induction, to which brief reference will be made at the end of this chapter.

3 | Theoretical Support

The support that may be claimed for a hypothesis need not all be of the inductive-evidential kind that we have considered so far: it need not consist entirely—or even partly—of data that bear out test implications derived from it. Support may also come "from above"; that is, from more inclusive hypotheses or theories that imply the given one and have independent evidential support. To illustrate: [consider the] hypothetical law for free fall on the moon, $s = 2.7 t^2$. Although none of its test implications have ever been checked by experiments on the moon, it has strong *theoretical support*, for it follows deductively from Newton's theory of gravitation and of motion (strongly supported by a highly diversified body of evidence) in conjunction with the information that the radius and the mass of the moon are .272 and .0123 of those of the earth and that the gravitational

acceleration near the surface of the earth is 32.2 feet per second per second.*

Similarly, the confirmation of a hypothesis that does have inductive-evidential support will be further strengthened if, in addition, it acquires deductive support from above. This happened, for example, to Balmer's formula. Balmer had anticipated the possibility that the hydrogen spectrum might contain further series of lines, and that the wavelengths of all the lines might conform to a generalization of his formula; namely,

$$\lambda = b \frac{n^2}{n^2 - m^2}$$

Here, m is a positive integer, and n is any integer greater than m . For $m = 2$, this generalization yields Balmer's formula; whereas $m = 1, 3, 4, \dots$ determine new series of lines. And indeed, the existence of the series corresponding to $m = 1, 3, 4$, and 5 was later established by experimental exploration of the invisible infrared and ultraviolet parts of the hydrogen spectrum. Thus, there was strong evidential support for a more general hypothesis that implied Balmer's original formula as a special case, thus providing deductive support for it. And deductive support by a theory came in 1913, when the generalized formula—hence Balmer's original one, also—were shown by Bohr to be derivable from his theory of the hydrogen atom. This derivation greatly strengthened the support of Balmer's formula by fitting it into the context of quantum-theoretical conceptions developed by Planck, Einstein, and Bohr, which were supported by diverse evidence other than the spectroscopic measurements that lent inductive support to Balmer's formula.³

Correlatively, the credibility of a hypothesis will be adversely affected if it conflicts with hypotheses or theories that are accepted at the time as well-confirmed. In the *New York Medical Record* for 1877, a Dr. Caldwell of Iowa, reporting on an exhumation he claims to have witnessed, asserts that the hair and the beard of a man who had been buried clean-shaven, had burst the coffin and grown through the cracks.⁴ Although presented by a presumptive eyewitness, this statement will be rejected without much hesitation because it conflicts with well-established findings about the extent to which human hair continues to grow after death. . . .

The principle here referred to [that conflict with a broadly supported theory militates against a hypothesis] must be applied with discretion and

* Hempel wrote this in 1966, three years before the first moon landing by Apollo 11 astronauts Neil Armstrong and Edwin ("Buzz") Aldrin. Newton's theory entails that a , the acceleration due to gravity at the surface of any planet, is directly proportional to the planet's mass and inversely proportional to the square of its radius. For any planet of reasonable size, a will be nearly constant close to the planet's surface, and simple algebra yields the distance formula $s = \frac{1}{2}at^2$ for a falling body dropped from rest.

restraint, however. Otherwise, it could be used to protect any accepted theory against overthrow: adverse findings could always be dismissed as conflicting with a well-established theory. Science does not, of course, follow this procedure; it is not interested in defending certain pet conceptions against all possible adverse evidence. It aims, rather, at a comprehensive body of sound empirical knowledge, represented by a well-confirmed system of empirical statements, and it is accordingly prepared to give up or to modify whatever hypotheses it may have previously accepted. But findings that are to dislodge a well-established theory have to be weighty; and adverse experimental results, in particular, have to be repeatable. Even when a strong and useful theory has been found to conflict with an experimentally reproducible "effect", it may still continue to be used in contexts where it is not expected to lead into difficulties. For example, when Einstein propounded the theory of light quanta to account for such phenomena as the photoelectric effect, he noted that in dealing with the reflection, refraction, and polarization of light, the electromagnetic wave theory would probably never be replaced; and it is indeed still used in this context. A large-scale theory that has been successful in many areas will normally be abandoned only when a more satisfactory alternative theory is available—and good theories are difficult to come by.⁵

4 | Simplicity

Another aspect that affects the acceptability of a hypothesis is its simplicity, compared with that of alternative hypotheses that would account for the same phenomena.

Consider a schematic illustration. Suppose that investigation of physical systems of a certain type (Cepheids, elastic metal springs, viscous liquids, or whatever) suggests to us that a certain quantitative characteristic, v , of such systems, might be a function of, and thus uniquely determined by, another such characteristic, u (in the way in which the period of a pendulum is a function of its length). We therefore try to construct a hypothesis stating the exact mathematical form of the function. We have been able to check many instances in which u had one of the values 0, 1, 2, or 3; the associated values of v were regularly found to be 2, 3, 4, and 5, respectively. Suppose further that concerning these systems, we have no background knowledge that might bear on the likely form of the functional connection, and that the following three hypotheses have been proposed on the basis of our data:

$$H_1: v = u^4 - 6u^3 + 11u^2 - 5u + 2$$

$$H_2: v = u^5 - 4u^4 - u^3 + 16u^2 - 11u + 2$$

$$H_3: v = u + 2$$

Each of these fits the data: to each of the four u -values examined, it assigns exactly the v -value that has been found associated with it. In geometrical terms: if the three hypotheses are graphed in a plane coordinate system, then each of the resulting curves contains the four data-points $(0,2)$, $(1,3)$, $(2,4)$, and $(3,5)$.

Yet if, as has been assumed, we have no relevant background information that might indicate a different choice, we would no doubt favor H_3 over H_1 and H_2 on the ground that it is a simpler hypothesis than its rivals. This consideration suggests that if two hypotheses accord with the same data and do not differ in other respects relevant to their confirmation, the simpler one will count as more acceptable.

The relevance of the same basic idea to entire theories is often illustrated by reference to the Copernican heliocentric conception of the solar system, which was considerably simpler than the geocentric one it came to supersede, namely, Ptolemy's ingenious and accurate, but "gorgeously complicated system of main circles and sub-circles, with different radii, speeds, tilts, and different amounts and directions of eccentricity."⁶

Though, undeniably, simplicity is highly prized in science, it is not easy to state clear criteria of simplicity in the relevant sense and to justify the preference given to simpler hypotheses and theories.

Any criteria of simplicity would have to be objective, of course; they could not just refer to intuitive appeal or to the ease with which a hypothesis or theory can be understood or remembered, etc., for these factors vary from person to person. In the case of quantitative hypotheses like H_1 , H_2 , H_3 , one might think of judging simplicity by reference to the corresponding graphs. In rectangular coordinates, the graph of H_3 is a straight line, whereas graphs of H_1 and H_2 are much more complicated curves through the four data-points. But this criterion seems arbitrary. For if the hypotheses are represented in polar coordinates, with u as the direction angle and v as the radius vector, then H_3 determines a spiral, whereas a function determining a "simple" straight line would be quite complicated.

When, as in our example, all the functions are expressed by polynomials, the order of the polynomial might serve as an index of complexity; thus H_2 would be more complex than H_1 , which in turn would be more complex than H_3 . But further criteria are needed when trigonometric and other functions are to be considered as well.

In the case of theories, the number of independent *basic assumptions* is sometimes suggested as an indicator of complexity. But assumptions can be combined and split up in many ways: there is no unambiguous way of counting them. For example, the statement that for any two points there is exactly one straight line containing them might be counted as expressing two assumptions rather than one: that there is at least one such line, and that there is at most one. And even if we could agree on the count, different basic assumptions might in turn differ in complexity and would then have to be weighed rather than counted. Similar remarks apply to

the suggestion that the number of *basic concepts* used in a theory might serve as an index of its complexity. The question of criteria of simplicity has in recent years received a good deal of attention from logicians and philosophers, and some interesting results have been obtained, but no satisfactory general characterization of simplicity is available. As our examples suggest, however, there certainly are cases in which, even in the absence of explicit criteria, investigators would be in substantial agreement about which of two competing hypotheses or theories is the simpler.

Another intriguing problem concerning simplicity is that of justification: what reasons are there for following the *principle of simplicity*, as we might call it, that is, the maxim that the simpler of two otherwise equally confirmed rival hypotheses or theories is to be preferred, is to count as more acceptable?

Many great scientists have expressed the conviction that the basic laws of nature are simple. If this were known, there would indeed be a presumption that the simpler of two rival hypotheses is more likely to be true. But the assumption that the basic laws of nature are simple is of course at least as problematic as the soundness of the principle of simplicity and thus cannot provide a justification for it.

Some scientists and philosophers—among them Mach, Avenarius, Ostwald, and Pearson—have held that science seeks to give an economic or parsimonious description of the world, and that general hypotheses purporting to express laws of nature are economic expedients for thought, serving to compress an indefinite number of particular cases (e.g., many cases of free fall) into one simple formula (e.g., Galileo's law); and from this point of view, it seems entirely reasonable to adopt the simplest among several competing hypotheses. This argument would be convincing if we had to choose between different *descriptions of one and the same set of facts*; but in adopting one among several competing hypotheses, such as H_1 , H_2 , H_3 above, we also adopt the *predictions* it implies concerning as yet untested cases; and in this respect, the hypotheses differ widely. Thus, for $u = 4$, H_1 , H_2 , and H_3 predict the v -values 150, 30, and 6, respectively. Now, H_3 may be mathematically simpler than its rivals, but what grounds are there for considering it more likely to be true, for basing our expectations concerning the as yet unexamined case $u = 4$ on H_3 rather than on one of the competing hypotheses, which fit the given data with the same precision?

One interesting answer has been suggested by Reichenbach.⁷ Briefly, he argues as follows: suppose that in our example v is indeed a function of u , $v = f(u)$. Let g be its graph in some system of coordinates; the choice is inessential. The true function f and its graph are, of course, unknown to the scientist who measures associated values of the two variables. Assuming, for the sake of the argument, that his measurements are exact, he will thus find a number of data-points that lie on the "true" curve g . Suppose now that in accordance with the principle of simplicity, the sci-

entist draws the simplest, i.e., the intuitively smoothest, curve through those points. Then his graph, say g_1 , may deviate considerably from the true curve, though it does share at least the measured data-points with the latter. But as the scientist determines more and more data-points and plots further simplest graphs, g_2, g_3, g_4, \dots , these will coincide more and more nearly with the true curve g , and the associated functions of f_2, f_3, f_4, \dots will approximate more and more closely the true functional connection f . Thus, observance of the principle of simplicity cannot be *guaranteed* to yield the function f in one step or even in many; but if there is a functional connection between u and v , the procedure will gradually lead to a function that approximates the true one to any desired degree.

Reichenbach's argument, which has here been stated in a somewhat simplified form, is ingenious; but its force is limited. For no matter how far the construction of successive graphs and functions may have gone, the procedure affords no indication at all of how close an approximation to the true function has been attained—if indeed there is a true function at all. (As we noted earlier, for example, the volume of a body of gas may seem to be, but is not in fact, a function of its temperature alone.) Moreover, the argument on grounds of convergence towards the true curve could be used also to justify certain other, intuitively complex and unreasonable methods of plotting graphs. For example, it is readily seen that if we were always to connect any two adjacent data-points by a semicircle whose diameter is the distance between the points, the resulting curves would eventually converge toward the true curve if there is one. Yet despite this "justification", this procedure would not be regarded as a sound way of forming quantitative hypotheses. Certain other nonsimple procedures, however—such as connecting adjacent data-points by hairpin loops whose length always exceeds a specified minimum value—are not justifiable in this fashion and can indeed be shown by Reichenbach's argument to be self-defeating. His idea is thus of distinct interest.

A very different view has been advanced by Popper. He construes the simpler of two hypotheses as the one that has greater empirical content, and he argues that the simpler hypothesis can therefore more readily be falsified (found out to be false), if indeed it should be false; and that this is of great importance to science, which seeks to expose its conjectures to the most thorough test and possible falsification. He summarizes his argument as follows: "Simple statements, if knowledge is our object, are to be prized more highly than less simple ones *because they tell us more, because their empirical content is greater; and because they are better testable.*"⁸ Popper makes his notion of degree of simplicity as degree of falsifiability more explicit by means of two different criteria. According to one of them, the hypothesis that the orbit of a given planet is a circle is simpler than the hypothesis that it is an ellipse, because the former could be falsified by the determination of four positions that are found not to lie on a circle (three positions can always be connected by a circle),

whereas the falsification of the second hypothesis would require the determination of at least six positions of the planet. In this sense, the simpler hypothesis is here the more readily falsifiable one, and it is also stronger because it logically implies the less simple hypothesis. This criterion surely contributes to clarifying the kind of simplicity that is of concern to science.

But Popper alternatively calls one hypothesis more falsifiable, and hence simpler, than another if the first implies the second and thus has greater content in a strictly deductive sense. However, greater content is surely not always linked to greater simplicity. To be sure, sometimes a strong theory, such as Newton's theory of gravitation and motion, will be regarded as simpler than a vast array of unrelated laws of more limited scope that are implied by it. But the desirable kind of simplification thus achieved by a theory is not just a matter of increased content; for if two unrelated hypotheses (e.g., Hooke's and Snell's laws) are conjoined, the resulting conjunction tells us more, yet is not simpler, than either component. Also, of the three hypotheses H_1 , H_2 , H_3 considered above, none tells us more than any of the others; yet they do not count as equally simple. Nor do those three hypotheses differ in point of falsifiability. If false, any one of them can be shown to be false with the same ease—namely, by means of one counter-instance; for example, the data-pair (4, 10) would falsify them all.

Thus, while all the different ideas here briefly surveyed shed some light on the rationale of the principle of simplicity, the problems of finding a precise formulation and a unified justification for it are not as yet satisfactorily solved.⁹

5 | The Probability of Hypotheses

Our survey of factors determining the credibility of scientific hypotheses shows that the credibility of a hypothesis H at a given time depends, strictly speaking, on the relevant parts of the total scientific knowledge at that time, including all the evidence relevant to the hypothesis and all the hypotheses and theories then accepted that have any bearing upon it; for as we have seen, it is by reference to these that the credibility of H has to be assessed. Strictly, therefore, we should speak of the *credibility of a hypothesis relative to a given body of knowledge*; the latter might be represented by a large set K of statements—all the statements accepted by science at the time.

The question naturally suggests itself whether it is possible to express this credibility in precise quantitative terms, by formulating a definition which, for any hypothesis H and any set K of statements, determines a number $c(H, K)$ expressing the degree of credibility that H possesses relative to K . And since we often speak of hypotheses as more or less probable,

we might wonder further whether this quantitative concept could not be so defined as to satisfy all the basic principles of probability theory. In this case, the credibility of a hypothesis relative to any set K would be a real number no less than 0 and no greater than 1, a hypothesis that is true on purely logical grounds (such as 'Tomorrow it will rain in Central Park or it won't') would always have the credibility 1, and finally, for any two logically incompatible statements H_1 and H_2 , the credibility of the hypothesis that one or the other of them is true would equal the sum of their credibilities $c(H_1 \text{ or } H_2, K) = c(H_1, K) + c(H_2, K)$.

Various theories for such probabilities have indeed been proposed.¹⁰ They proceed from certain axioms like those just mentioned to a variety of more or less complex theorems that make it possible to determine certain probabilities *provided that others are already known*; but they offer no general definition of the probability of a hypothesis relative to given information.

And if the definition of the concept $c(H, K)$ is to take account of all the different factors we have surveyed, then the task is very difficult, to say the least; for as we saw, it is not even clear how such factors as the simplicity of a hypothesis, or the variety of its supporting evidence, are to be precisely characterized, let alone expressed in numerical terms.

However, certain illuminating and quite far-reaching results have recently been obtained by Carnap, who has studied the problem by reference to rigorously formalized model languages whose logical structure is considerably simpler than that required for the purpose of science. Carnap has developed a general method of defining what he calls the degree of confirmation for any hypothesis expressed in such a language with respect to any body of information expressed in the same language. The concept thus defined does satisfy all the principles of probability theory, and Carnap accordingly refers to it as the *logical* or *inductive probability* of the hypothesis relative to the given information.¹¹

■ | Notes

1. W. F. Magie, ed., *A Source Book in Physics*, p. 74.
2. A full and lucid account, on which this brief survey is based, will be found in Chap. 33 of G. Holton and D. H. D. Roller, *Foundations of Modern Physical Science* (Reading, Mass.: Addison-Wesley Publishing Co., 1958).
3. For details, see Holton and Roller, *Foundations of Modern Physical Science*, Chap. 34 (especially section 7).
4. B. Evans, *The Natural History of Nonsense* (New York: Alfred A. Knopf, 1946), p. 133.
5. This point is suggestively presented and illustrated by reference to the phlogiston theory of combustion in Chap. 7 of J. B. Conant, *Science and Common Sense*.

A provocative general conception of the rise and fall of scientific theories is developed in T. S. Kuhn's book *The Structure of Scientific Revolutions* (Chicago: The University of Chicago Press, 1962)

6 E. Rogers, *Physics for the Inquiring Mind* (Princeton: Princeton University Press, 1960), p. 240. Chapters 14 and 16 of this work offer a splendid description and appraisal of the two systems; they give more substance to the claim of greater simplicity for Copernicus' scheme, but show also that it was able to account for various facts, known at Copernicus' time, that the Ptolemaic system could not explain

7 H. Reichenbach, *Experience and Prediction* (Chicago: The University of Chicago Press, 1938), section 42.

8 K. R. Popper, *The Logic of Scientific Discovery* (London: Hutchinson, 1959), p. 142 (italics are quoted). Chapters VI and VII of this book, which offer many illuminating observations on the role of simplicity in science, contain the presentation of the ideas here referred to.

9 The reader who wishes to pursue these issues further will find the following discussions helpful: S. Barker, *Induction and Hypothesis* (Ithaca: Cornell University Press, 1957); "A Panel Discussion of Simplicity of Scientific Theories," *Philosophy of Science*, Vol. 28 (1961), 109-71; W.V.O. Quine, "On Simple Theories of a Complex World," *Synthese*, Vol. 15 (1963), 103-6.

10 One of them by the economist John Maynard Keynes, in his book, *A Treatise on Probability* (London: Macmillan & Company, Ltd., 1921).

11 Carnap has given a brief and elementary account of the basic ideas in his article "Statistical and Inductive Probability," reprinted in E. H. Madden, ed., *The Structure of Scientific Thought* (Boston: Houghton Mifflin Company, 1960), pp. 269-79. A more recent, very illuminating statement is given in Carnap's article, "The Aim of Inductive Logic" in E. Nagel, P. Suppes, and A. Tarski, eds., *Logic, Methodology and Philosophy of Science*. Proceedings of the 1960 International Congress (Stanford: Stanford University Press, 1962), pp. 303-18.

Is Evidence Historical?

Suppose that some startling new phenomenon is observed. It is unlike anything which has been observed before, and for quite some time the scientific community cannot explain it. Then a new theory is proposed which is able to explain the phenomenon. Should we thereby consider this theory to have evidence in its favor? That is, should we count the fact that the theory provides an explanation for this phenomenon as evidence that the theory is true?

Many philosophers of science would say no. A common view is that explanations do *not* count as evidence for a theory's truth, only successful predictions of new phenomena do. If the new theory had predicted the occurrence of this startling phenomenon *before* it occurred, then (once it did occur) the theory would have evidence in its favor. Recently, however, this position requiring predictions for evidence has been challenged by a historian of science who claims that only explanations of previously known facts—and not predictions—can count as evidence for a theory.

Both sides in the dispute seem supported by some common intuitions. Predictions of new phenomena are extremely impressive, and seem to endow scientists with a kind of mystical power to foretell the future. At the same time, the view requiring that a theory explain some known fact which other theories have been unable to explain is also intuitively appealing; we may feel that a theory's success where others have failed must indicate something about its truth. In this paper I will put these intuitions to the test.

Before examining these opposing positions, I must clarify the terms being used by them. "Prediction" does not necessarily refer to a future event, as it does in our ordinary usage of the term, but to a correct state-

ment about some presently unknown phenomenon, whether it is a past, present, or future event. Examples include predicting that the extinction of the dinosaurs was caused by a collision on earth of an enormous meteor; that there are six quarks; that the ozone layer will be dangerously depleted in 100 years. Each of these facts is not yet known to be true; once discovered to be true, the prediction would be considered successful. On the other hand, "explanation" as the term is used here has to do with facts already known to be true. In this dispute, a theory is said to offer an explanation when it can account for a known fact. For example, Newton's theory of universal gravitation offered an explanation for the previously known fact that the tides ebb and flow.

So the dispute over whether evidence must consist in explanations or in predictions is actually concerned with the issue of when the evidence is known to be true relative to the time that the theory is proposed. If information *e* is known when theory *T* is proposed, and *T* explains *e*, then *e* is "old evidence." If *T* entails *e*, but *e* is not known when *T* is proposed, then *e* is "new evidence." In this paper I examine both the claim that evidence must always consist in new evidence, and the contrasting claim that evidence must always consist in old evidence. I argue that both positions are untenable, because the time at which *e* is known to be true relative to the theory's invention is not relevant to its status as evidence. Hence I reject what has been called the "historical thesis" of evidence, which holds that the time at which *e* is known is relevant. Evidence can consist in explanations as well as in predictions. This claim is clarified by an examination of the concept of evidence, and the introduction of a new theory of scientific evidence.

■ | The Preference for Prediction

In this section I will briefly outline three types of arguments for the view that only successful predictions can count as evidence for a theory.

THE "NO COINCIDENCE" ARGUMENT

In 1846, the planet Neptune was discovered after its existence and position had been predicted independently by two men, U. J. J. LeVerrier and John Couch Adams. Perturbations in the orbit of Uranus expected on Newtonian theory had led these men to conclude that there must be an unobserved body external to Uranus' orbit exerting an additional gravitational force on the planet. Using Newtonian theory, they were able to calculate mathematically the mass and orbit of this unseen planet. Acting upon LeVerrier's calculations, astronomers at the Berlin Observatory found the planet less than one degree from its expected location. From

Newton's theory, it was possible to predict successfully the existence, position, and mass of a previously unexpected planet. This success was considered further evidence for Newton's theory of universal gravitation.

William Whewell, like other nineteenth-century intellectuals, praised the discovery of Neptune as a triumph of astronomy, which he termed the "Queen of the Sciences." Astronomy deserved this royal title because of its success at predicting phenomena: not only new instances of known phenomena, such as eclipses, but also novel and previously unexpected phenomena such as an eighth planet. Whewell argued—most notably in a published interchange with John Stuart Mill—that "to predict unknown facts found afterwards to be true is . . . a confirmation of a theory which in impressiveness and value goes beyond any explanation of known facts" (1857, 557).

Whewell's position (unlike the others I will discuss in this section) is that while explanations can count as some evidence for a theory, predictions are much stronger evidence. I will concentrate here on the argument for the evidential value of predictions, because Whewell's type of argument can be used to deny any evidential value to explanations at all. The reasoning behind this position is as follows. Whewell claims that the agreement of the prediction with what occurs (that is, the fact that the prediction turns out to be correct) is "nothing strange, if the theory be true, but quite unaccountable, if it be not" (1860, 273–4). For example, if Newtonian theory were not true, Whewell would argue, the fact that from the theory we could correctly predict the existence, location, and mass of Neptune would be bewildering, and indeed miraculous. He is using what Gilbert Harman (1965) calls the "inference to the best explanation" argument. That is, Whewell is claiming that the best explanation for the success of the prediction is that the theory which made the prediction is true. Further, he is claiming that because it is the "best" explanation, it is the "correct" one. Whewell is not claiming that there is no other logically possible explanation for the agreement between prediction and event; he admits that this agreement might be due to "mere chance" or "coincidence." But Whewell rejects this possible explanation as clearly inferior to the explanation that the theory is true.¹

On the other hand, the mere fact that a theory can explain already known phenomena does not count as (such strong) evidence for a theory's truth. If a theory is invented for the express purpose of explaining some phenomenon, the fact that it does explain it is certainly no coincidence. The reasoning seems to be that, when a theory can explain some known phenomenon, the possible reasons for the theory's explanatory success are not merely either "coincidence" or "truth": another possibility is added—the "ingenuity" of the theory's inventor. Since the theorist knew what was needed to make the theory "successful" (that is, it needed to explain a certain known fact), it is not surprising that the theory should be able to do so. So the "best explanation" for the theory's success is no longer that

the theory is true; rather, it is that the theory's inventor exercised much ingenuity in engineering a theory in order to explain the phenomenon. Whewell shares the suspicion with many contemporary writers that a theorist who knows the facts that need explaining is somehow inappropriately influenced by knowing what the theory has to do. Those who require predictions to rule out such "ingenuity" are suggesting that predictive successes—unlike successful explanations—are (as Whewell puts it) "beyond the stamp of ingenuity to counterfeit."²

THE FALSIFICATION ARGUMENT

In distinguishing science from "pseudosciences" such as astrology, people often point to predictive capability. The power to make predictions, which are then confirmed, is considered by many to be the hallmark of scientific activity. This view, and the fact that it is so widespread among the general public, are due in great measure to the writings of the philosopher Karl Popper. According to Popper, science aims at "bold conjectures" about the world. Proper scientific method consists in inventing such conjectures, or theories, and then attempting to find evidence which proves them to be false. No evidence can ever prove a theory to be true, or even probably true; but a theory which has withstood the sincere attempt at being proven false (that is, which has not been "falsified") is considered to be "highly corroborated." The best that any evidence can do, then, is to corroborate a theory.

On this view, only predictions can count as evidence, because only predictions are "potential falsifiers" of a theory. If a theory explains a known fact, this information cannot potentially falsify the theory because it is already known that the fact explained is true. But if a theory predicts an unknown fact (and if this prediction is verifiable observationally) then it can function as a potential falsifier: we can attempt to observe whether this phenomenon does, in fact, occur. If it does not, then we have falsified the theory which predicted it. If the predicted phenomenon *does* occur, then the theory is corroborated—that is, tested and not yet falsified.³

For example, according to this view, the fact that Kepler's elliptical theory of Mars' orbit could explain Tycho Brahe's data is not evidence for the theory. Because Brahe's data were known, the fact that Kepler could explain them was not potentially falsifying for his theory. That is, the data did not provide a means for testing and possibly rejecting the theory. However, Kepler's general theory that the orbits of all the planets describe ellipses could be falsified. This theory entails predictions—for example, about as yet unobserved planets—that could be tested in order to falsify the theory. When Uranus was discovered in 1781, and Neptune in 1846, their orbits were found to describe ellipses; in this way, Kepler's theory received corroboration.

THE POSITIVE-RELEVANCE ARGUMENT

Many recent accounts of evidence are quantitative, and use probability theory in order to explicate the concept. One type of quantitative account claims that for some piece of information to be evidence for a hypothesis, the information must raise the probability that the hypothesis is true. Some, following Rudolf Carnap, call this the "positive relevance" view of evidence; it is also referred to as the "increase in probability" view.

On this view, if h is the hypothesis and e is the putative evidence,

e is evidence for h if and only if
 $p(h/e) > p(h)$.

That is, e is evidence for h if and only if the probability of h given e is greater than the probability of h alone. The increase in probability view is generally taken to entail that only predictions can be evidence for a hypothesis. This can be demonstrated using Bayes' theorem, which follows from the probability calculus.

$$\text{Bayes' theorem: } p(h/e) = \frac{p(h) \times p(e/h)}{p(e)}$$

It is claimed that, by definition, a known fact has a probability equal to one (it is "certain"). That is, $p(e) = 1$. Now if $p(e) = 1$, then, because h is consistent with e (because h explains e), $p(e/h) = 1$. Hence, using Bayes' theorem,

$$p(h/e) = p(h).$$

That is, e does *not* increase the probability that h is true when e is a known fact that h explains. On this positive relevance view of evidence, then, the explanation of a known fact cannot be considered evidence for a theory.

To illustrate this view, let e be Brahe's observations and h be the hypothesis that the orbit of Mars describes an ellipse. Since Brahe's observations are a known fact, $p(e) = 1$; and since h is consistent with e , $p(e/h) = 1$. Using Bayes' theorem, we get the result that the probability that h is true given e is equal to the probability that h is true without e . Hence, e does not raise h 's probability, and so it is not evidence for h .

A successful prediction, on the other hand, always counts as evidence according to this view. Because a prediction is not yet known to be true (it is not certain), $p(e) < 1$ when e is a prediction. And since h entails e (that is, h leads to the prediction that e), $p(e/h) = 1$. Again using Bayes'

theorem, we get

$$p(h/e) > p(h).$$

Kepler's hypothesis that the orbit of Mars describes an ellipse entails the prediction e' that tomorrow, Mars will be in a certain position. Since it is at present uncertain that e' is true (since it has not yet occurred), $p(e') < 1$. However, since h entails e' , $p(e'/h) = 1$. Therefore when e' is found to be true, it will raise the probability that h is true, since $p(h/e') > p(h)$.⁴

■ | The Case for Explanation

In a recent series of articles the historian of science Stephen Brush has argued against all views requiring that evidence be "new."⁵ His claim is that a prediction cannot be "reliable" evidence for a theory; only explanations of known facts can count as evidence for scientific theories.⁶ Brush illustrates his claim with several examples, including one drawn from general relativity theory. I will briefly examine this example, in order to explicate his position.

After Einstein devised his theory of general relativity, he found that it could offer an explanation of a previously known fact: it could account for the advance of Mercury's perihelion. This phenomenon (referred to as "Mercury's orbit") is the fact that Mercury is closest to the sun at different points of its orbit during successive revolutions around the sun. The presence of other planets causes the orbit of a planet to be perturbed—that is, the axis of the orbit's ellipse rotates slowly relative to the fixed stars. In the case of Mercury, the observed perturbation is larger than that which is calculated by Newtonian theory. Einstein was able to construct an elaborate derivation showing that general relativity theory entailed this discrepancy of 43" of arc per century between the theoretical (Newtonian) value and the observed position. Einstein showed that if the initial orbit of Mercury is calculated using general relativity theory, and the value thus obtained is added to the Newtonian value of perturbation caused by other planets, this total exactly conforms to the observed result. Hence his new theory was able to explain a known fact previously considered inexplicable.

Einstein's theory also entailed two predictions of novel phenomena. One of these was the gravitational redshift of spectral lines (that is, the prediction that the frequencies of celestial spectral lines should be shifted to measurably lower frequencies compared to those of terrestrial spectral lines). This prediction was not accurately confirmed experimentally until 1960. The other, which was confirmed during Einstein's lifetime, was the phenomenon of light-bending. Einstein's theory entailed the prediction that light passing a massive body such as the sun on its way to Earth would

be attracted by the body's gravitational force. As a result, light from a distant star should be deflected as it passes the sun on its way to the Earth. In 1915 Einstein made his final calculation of the degree of deflection which should occur. This result could be observed by comparing the relative positions of a star when the sun is near the path of its light and when the sun is not. The first position would be visible only during a solar eclipse, when the sun's own light would not be blocking the light from the nearby star. Einstein's prediction was confirmed during the solar eclipse of May 1919.

Brush claims that, contrary to what proponents of the predictivist view would argue, this prediction of a quite unexpected phenomenon was not better evidence than the explanation of Mercury's orbit; rather, the explanation was better evidence for the truth of general relativity theory. His reasoning is that

a successful explanation of a fact that other theories have already failed to explain satisfactorily (for example, the Mercury perihelion) is more convincing than the prediction of a new fact, *at least until competing theories have had a chance* (and failed) to explain it. (1989, 1127, my emphasis)

By the time Einstein gave his explanation for Mercury's orbit, other theories had tried and failed to explain it. However, since light-bending was a newly discovered phenomenon, other competing theories had not yet had the opportunity to explain it. Besides claiming that the explanation was "more convincing," Brush makes the stronger, noncomparative claim that a successful prediction cannot become good or "reliable" evidence for a theory until other theories have had their shot at explaining the new phenomenon. (I call this the "equal opportunity" requirement of Brush's view.) He asserts that "light-bending could not become reliable evidence for Einstein's theory until those alternatives failed" (1989, 1127). Hence, Brush claims "it was only ten years after the initial report of light-bending" that it could be "plausibly asserted" that the phenomenon was evidence for Einstein's theory (1989, 1126).⁷

■ | The Concept of Evidence

All of the views which claim either that evidence must consist in predictions of unknown phenomena or that it must consist in explanations of known facts agree with what has been called the "historical thesis."⁸ This thesis holds that, on any plausible theory of evidence, the time at which *e* is known relative to the invention of *h* is relevant to whether *e* is evidence for *h*. In the case of the predictivist positions examined, *e* is evidence for *h* only if *e* was not known to be true at the time that *h* was proposed, but was discovered to obtain sometime thereafter. Brush's view

also entails the historical thesis, by claiming that for *e* to be evidence for *h*, *e* *must* have been known to be true before *h* was proposed. Other views which conform to the general historical thesis are those which claim that the time at which *e* is known to be true is relevant to how *strongly* *e* confirms *h*: for example, Whewell's view that predictions are stronger evidence than explanations. In the remainder of this paper I will argue that the historical thesis is incompatible with any theory of evidence that captures how the concept is used in science. Hence both positions in the explanation versus prediction dispute are untenable.

What is evidence? That is, what is it to claim that some information *e* is evidence for, or confirms, a hypothesis *h*? Roughly speaking, it is to claim that *e* gives a reason for believing in the truth of *h*.⁹ When we say that the spots on John's face are evidence that he has the measles, we mean that the presence of the spots gives a reason to believe that John has the measles. Note that evidence need not be conclusive.¹⁰ There can be a reason to believe that John has the measles without it being true that he is so afflicted.

Reasons for belief can be either personal or impersonal. Personal reasons for belief are relativized to a particular person; that is, some information *e* might be person *P*'s reason for believing a hypothesis. There are two ways in which a reason for belief can be personal. The first is *factual*: *e* is person *P*'s *factual* reason for believing *h* only if *P* in fact believes *h* for the reason that *e*. There is also a *normative* way in which a reason for belief can be personal: *e* is person *P*'s *normative* reason for believing *h* only if, given *P*'s knowledge and beliefs, *e* gives a reason for *P* to believe *h*, whether or not *P* does believe *h* for the reason that *e*. Another way of expressing this normative relativization is that person *P* *should* believe *h* for the reason that *e*, even if, in fact, *P* does not believe *h* at all or believes *h* for another reason.

A reason for belief that is impersonal is one which is not relativized to any particular person. That is, if *e* is an *impersonal* reason to believe *h*, then it provides a reason for anyone to believe *h*, whether or not any particular person does, in fact, believe *h* for the reason that *e*, or whether any person should, given his other beliefs and knowledge, believe *h* for the reason that *e*. For example, the fact that the water in the puddle outside is frozen solid (*e*) is an impersonal reason for anyone to believe that the outside temperature is zero degrees Celsius or below (*h*), even if a particular person does not (and should not, given his lack of knowledge of the Celsius scale) believe *h* because of *e*.

In science, evidence is used in the *impersonal* sense. When a scientist publishes experimental results, claiming that these constitute evidence for a certain theory, she is not claiming merely that these results constitute her personal (factual or normative) reason for believing the theory. Rather, the scientist is claiming that the results constitute a reason—a reason for anyone—to believe the theory. Of course, others may disagree with this

assessment; but in such cases the disagreement is over whether some information truly is a reason for everyone to believe *h*, or whether it is reason enough.

This impersonal use of evidence is seen in several professional activities performed by scientists. First, scientists engage in public debate over theory-choice. That is, they argue amongst themselves over which theory of a group of competitors is the most reasonable theory to believe. For example, some medical researchers are currently arguing over whether there is more reason to believe that HIV infection *causes* AIDS, or that it is not a cause but merely a coexisting condition.¹¹ Scientists generally invoke evidence in these debates (for example, the evidence that a small number of patients with AIDS do not seem to be infected with HIV). The use of evidence in debating over theory-choice demonstrates that scientists do not consider evidence to be a personal concept. If scientific evidence were personal in the factual sense, then there would be no basis for *debate* at all; scientists S1 and S2 could merely *report* the hypotheses they each believe and the reasons for which they believe them.

It might be argued, against my claim, that argument over theory-choice takes place at the level of normative personal reasons for belief; that is, that when a scientist argues that *e* is a reason to believe *h* he means that other scientists who share his beliefs and knowledge should believe *h* because of *e*. While scientists may make this kind of argument, by so arguing they are not utilizing the normative personal concept of evidence. In such argument an appeal is made to beliefs and knowledge assumed to be shared by a group, not to the specific beliefs and knowledge of a particular individual. Moreover, a scientist invoking evidence in theory-choice debate is claiming *more* than just "those scientists who share my knowledge and beliefs ought to believe *h* because of *e*." The scientist is claiming, rather, that *e* is a reason for *anyone* to believe *h*, even those who lack his knowledge and beliefs. Thus, for example, the medical researchers arguing for the hypothesis that HIV is not the cause of AIDS are claiming that their evidence is a reason for *anyone*—other medical researchers, doctors, politicians, the general public—to believe *h*, no matter what other beliefs they might have. This is clearly a use of evidence as an impersonal concept. Of course, there is not always enough evidence known to choose between theories on the basis of the evidence; but my claim is that on any plausible theory of evidence, it must be possible in principle for such a choice to be made.¹²

Another professional activity performed by scientists is to evaluate, and occasionally to reject, the evidential claims of other scientists. This activity also exemplifies the impersonal character of evidence in science. For example, in recent years the scientific community has judged that the experiment which was claimed to provide evidence for cold fusion does not provide evidence that cold fusion occurred. Note that this judgment is not about whether the experimenters Fleischmann and Pons had a personal

reason (either factual or normative) to believe in cold fusion because of their experiment. Rather, scientists have concluded that the reported experimental data are not, in fact, reason for anyone to believe that cold fusion occurred. Fleischmann and Pons are therefore judged to have been incorrect in their claim that there is an impersonal reason to believe in cold fusion.*

Any plausible theory of evidence (that is, one which captures how the concept is used in science) must account for this impersonal use of evidence in science. Various philosophical explications of the concept of evidence have been proposed. These are either objective or nonobjective. A general condition for objective concepts of evidence is as follows:

Objective Concept of Evidence: whether e is evidence for h does not depend upon anyone's beliefs or knowledge about e , h , or anything else. Hence if some e is evidence for h , it is so regardless of what any person knows or believes.

Particular objective theories of evidence are given by Carnap, Hempel, Glymour, and Achinstein, among others.¹³ All satisfy this general objective condition. For example, Hempel's inductive account of evidence claims that a universal hypothesis is confirmed by its positive instances; hence the law "all crows are black" is confirmed by the existence of a black crow. This is so whether any person knows or believes that the crow exists, and even if no person knows or believes that a black crow is evidence for the hypothesis "all crows are black." Glymour's "bootstrapping" model of confirmation builds upon this positive-instance account, bringing in the idea that parts of a theory can be used to confirm other parts. In Glymour's

* The cold fusion fiasco began in March 1989, when Stanley Pons and Martin Fleischmann, two chemists working at the University of Utah, announced to the world's press that they had obtained controlled nuclear fusion at room temperature in a specially designed electrolytic cell containing heavy water. This announcement came before their work had been published in a refereed journal and while the details of their apparatus remained sketchy. Most physicists were skeptical because what Pons and Fleischmann described would violate widely accepted principles of nuclear physics. After some reports of confirmation by other groups (which were later retracted), repeated failures to duplicate the results of Pons and Fleischmann led everyone to the conclusion that there is no such thing as cold fusion. What Pons and Fleischmann had taken to be evidence for it—very small increases in temperature and the apparent production of neutrons—were artifacts of their equipment and measuring techniques. When rigorous controls were instituted, the effects disappeared. The case was controversial because, by going public the way that they did, Pons and Fleischmann circumvented the normal process of scientific review and publication. Such is the enormous commercial potential that cold fusion would have if it were real that large sums of money were earmarked for the research of Pons and Fleischmann solely on the basis of their premature announcement. The media hype surrounding it was intense but brief.

account, *e* confirms *h* (relative to some theory *T*) regardless of whether anyone knows or believes that *e* is true or that *e* is evidence for *h*. Achinstein's view incorporates the following conditions: *e* must be true, *h* must be highly probable given *e*, and it must be highly probable that there is an explanatory connection between *h* and *e*.¹⁴ Again, this view does not require that anyone knows or believes that *e* or *h* is true, or that any of these conditions obtain.

Objective theories of evidence have been preferred by many philosophers of science in part, I think, because any theory which satisfies this objective condition is able to capture the impersonal sense of evidence in science. (Note, however, that I am not claiming that *only* objective theories are impersonal; below I discuss a nonobjective impersonal theory of evidence.) Objective theories are not relativized to any person; when some information is deemed evidence for a hypothesis it is considered to be evidence for *h* for anyone. If the conditions of a particular objective theory are met, *e* is evidence for everyone. Conversely, if these conditions are not met, then *e* is not evidence for *h* for anyone. Thus objective theories can account for the fact that scientists invoke evidence as an impersonal concept in debating over theory-choice and in evaluating evidential claims.

Writers on the historical thesis have failed to realize that the historical thesis is incompatible with any theory of evidence that is objective. On any objective view, the *time* at which *e* is known relative to the invention of *h* cannot be relevant to whether *e* is evidence for *h*. As we have seen, the objective condition for evidence explicitly denies all relevance to what any person knows or believes about *e*, *h*, or anything else. So, for instance, if *e* is evidence for *h*, on an objective view, it is evidence for *h* whether or not any person knows or believes that *e* is true. On such views, if John has spots of a certain kind, then there is a reason to believe that he has measles, even if no one has seen these spots—that is, even if no person knows that *e* is true. Indeed on objective views of evidence, even if no person *ever* knows about the spots, they are still evidence that John has the measles. Clearly if *e* does not need to be known in order to confirm *h*, then it makes no sense to require that it must be known either before or after *h* is invented, in order to confirm it.

So if we want to accept the historical thesis, as many do, we must consider whether it is compatible with any *nonobjective* view of evidence, and whether any such view can capture the impersonal feature of scientific evidence. Defining “nonobjective” evidence in opposition to objective evidence, we get the following general condition:

Nonobjective Concept of Evidence: whether *e* is evidence for *h* at time *t* does depend on what someone (either a particular person or a particular group) knows or believes about *e*, *h*, and/or something else at *t*.

At first glance this approach looks promising in terms of the historical thesis, since nonobjective views consider what is known when the evidential claim is made to be relevant in determining whether *e* is evidence for *h*.

Let us first examine the most well-known nonobjective theory of evidence, "subjective Bayesianism." The subjective Bayesian view of evidence is a quantitative one, defining evidence in terms of probability (that is, as satisfying the standard axioms of the mathematical probability calculus). The subjective Bayesians are proponents of the positive-relevance view discussed in the first section of this paper; that is, they consider that evidence must raise the probability of a hypothesis. However, unlike objective probabilists, they define probability in subjective terms. According to them, the probability of *h* given *e* for a person *P* is defined as *P*'s degree of belief in *h* given the information expressed by *e*. A person's degrees of belief are characterized in terms of his attitudes towards real or potential bets on the truth of *h*. Subjective Bayesians demonstrate that one's set of degrees of belief in all hypotheses must satisfy the probability calculus in order to avoid the situation that one's betting behavior would necessarily result in loss. If one's set of degrees of belief satisfies the probability calculus, then subjective Bayesians speak of this set as *coherent*. Hence on this view, evidence can be defined in the following way: *e* is evidence for *h* for person *P* if and only if (i) *e* increases *P*'s degree of belief in *h*, and (ii) the degree of belief in *h* given *e* is part of a probabilistically coherent set of beliefs that *P* has.¹⁵

This nonobjective view of evidence is readily compatible with the historical thesis. In fact, prominent critics of subjective Bayesianism such as Glymour have rejected the view because of its supposed entailment of the predictivist side of the historical thesis.¹⁶ Recall that in the first section the positive-relevance view of prediction was discussed. There I demonstrated how, using Bayes' theorem, it was claimed that only predictions of unknown facts could raise the probability of a hypothesis. The subjective Bayesian view also requires an increase in probability. Hence the subjective Bayesian is taken to require predictions for evidence.¹⁷

Although subjective Bayesianism seems to entail the historical thesis, this theory does not capture the nature of evidence as it is used in science. In fact, the subjective Bayesian view denies that there is an impersonal concept of evidence; it makes evidence only a personal matter. Relative to their different sets of degrees of belief, two people can (and often must, to satisfy the coherence condition) have vastly different degrees of belief in a particular hypothesis given the same evidence.¹⁸ Indeed, the fact that John has spots of a certain kind on his face might be evidence for two contrary hypotheses. For John's doctor, the spots might be a reason to believe the hypothesis that John has the measles (and was not cursed by a witch). But at the same time, for John's superstitious grandfather the

spots might be evidence for the contrary hypothesis that John has been cursed by a witch. Because the same evidence can equally well support two conflicting hypotheses, relative to the differing sets of beliefs two people may possess, the subjective Bayesian view does not provide a basis for a scientific community to debate over theory-choice. On the basis of *e*, the doctor will choose one theory, the grandfather another, and neither will worry about the disagreement between them.

To a limited extent, the subjective Bayesian view does allow for the evaluation of evidential claims made by scientists. If scientist *P* claims that *e* is his evidence for *h*, but *P*'s degree of belief in *h* makes his complete set of beliefs incoherent, then we can reject *P*'s claim that *e* is his evidence for *h*. If the doctor claims that *e* is evidence that John has been cursed by a witch, we can reject this evidential claim as incoherent with the doctor's complete set of beliefs (assuming that it is incoherent with her beliefs). However, since such evaluation is relative to a particular person's set of degrees of belief, we cannot reject any evidential claim if it is consistent with the person's set of degrees of belief. So, for example, we cannot reject the grandfather's claim that John's spots are evidence that he has been cursed by a witch, as long as this belief is part of a probabilistically coherent set of beliefs held by the grandfather. On this view, scientists would not be justified in rejecting the evidential claims regarding cold fusion made by Fleischmann and Pons, if these claims are coherent with the experimenters' personal sets of beliefs.

Accordingly, subjective Bayesianism fails to capture the impersonal nature of evidence as the concept is used by scientists. However, I am not claiming that only objective notions of evidence can do so. On the contrary, there is a nonobjective view which captures this impersonal nature of scientific evidence, and which has not been discussed by philosophers of science. This is a more plausible nonobjective view than subjective Bayesianism. The concept of evidence it proposes is nonobjective because it depends on people's knowledge or beliefs; yet unlike the subjective Bayesian view, it is not relativized to the knowledge and beliefs of any particular person. Instead, it is relativized to some "total current body of scientific knowledge." On this view, evidence for a hypothesis *h* is a reason for any person to believe *h* at time *t*, given some body of knowledge known at *t*. Thus, for example, spots of a certain kind are evidence for measles at the present time, because (given our current medical knowledge), the presence of these spots on a patient's body is a reason for anyone to believe that the patient has measles.

By a "total body of knowledge" I mean what Newton means by "phenomena," as the term is used in the *Principia*.¹⁹ Newton suggests by his examples that the "phenomena" consist in facts that have been established by observation. Further, these facts are presumed to be noncontroversial; that is, the "phenomena" consist in facts that are agreed upon by scientists who are aware of the observations, and which would be agreed upon by

other scientists once they were made aware of them. Note, then, that not every person needs to know or believe these facts. Nonscientific laypeople may well not know all or even most of the facts constituting the "total body of knowledge." Even among scientists, there will be many who would not realize the truth of certain facts even if they were made aware of the relevant observations; for example, a medical researcher may not be in a position to agree or disagree on certain facts of physical astronomy. Here we can appeal, somewhat roughly, to the notion of "scientific experts" who establish the contents of the current body of scientific knowledge. Medical researchers establish the body of medical knowledge, while astronomers and physicists establish knowledge about celestial bodies and their motions. The important aspect of this notion is that the "body of knowledge" consists in noncontroversial facts accepted by the scientific experts (a weaker version would include beliefs which are generally thought to be facts by the experts but might later turn out to be false).

When there is disagreement among the experts in a particular field, as there is now over the causal relation between HIV and AIDS, the disputed information is not part of this body of accepted knowledge. Hence this is a *temporal* concept; what is considered accepted knowledge changes over time. For example, today it is accepted by the experts that the moon has mountains and valleys on its surface. But before Galileo turned the telescope to the heavens and observed the moon, the experts believed that the moon's surface was smooth. Even when Galileo observed that the moon had an irregular surface, this fact was controversial for many years, until it was accepted that the telescope did indeed accurately represent celestial bodies. Until the scientific experts agreed, this fact was not part of what I am calling the "total body of knowledge."²⁰

This nonobjective concept of evidence satisfies the following condition:

"Expert-relative" Concept: e is evidence for h at time t only if there is some body of information b that is known by the "experts" at t and, given b , e is a reason to believe h .²¹

This nonobjective concept of evidence, unlike the subjective Bayesian one, is able to account for the impersonal nature of evidence at any given time. On this view, evidence is not relativized to the knowledge and beliefs of any particular person; so if some e is determined to be evidence for h at time t , it is evidence for h (at t) for anyone. Unlike the objective views, this theory makes evidence a contextual matter, dependent upon the historical and epistemic circumstances surrounding the evidential claim. I will illustrate these aspects of the expert-relative concept of evidence by taking an intuitive (if somewhat gruesome) case. Suppose I discover a headless body with its severed head lying several feet away. The state of these body parts (e) constitutes very good reason for any person to believe

the hypothesis (h) "this person is dead," relative to all current medical knowledge (b). As opposed to the subjective Bayesian view of evidence, *e* is evidence for *h* for everyone, even for those people who do not know or believe this medical knowledge. And on this view, in contrast to the objective views, *e* might not always have been evidence for *h*. The first time a decapitated body was observed by a primitive society, it would not have been a reason to believe that the person was dead, if the "total current body of knowledge" did not yet include the information that humans cannot live without their heads attached to their bodies.

This expert-relative notion of evidence allows for the invoking of evidence as an impersonal concept in debates over theory-choice. In this example, *e* (John's decapitated body) provides a basis for debating over which hypothesis is more reasonable to believe at the present time: h_1 (John is dead) or h_2 (John is not dead). Given our current knowledge, *e* is a reason, at this time, for anyone to believe h_1 rather than h_2 . Further, on this nonobjective view it is possible to evaluate and reject evidential claims. And unlike on the subjective Bayesian view, these evaluations are not relative to any particular person's background beliefs or knowledge. On this view, *e* is evidence for *h* even if *e* is not my personal reason for believing in *h* (in either the factual or normative sense). For example, if I do not know that medical knowledge dictates that headless human bodies cannot be alive, or if I disagree with the experts on this, I might not believe *h* on the basis of *e*. Yet regardless of what I believe, *e* would still be an excellent reason now (given *b*) to believe that the person is dead. My claim that *e* is not evidence for *h* can—and should—be rejected.

On the expert-relative view, evidential claims must be evaluated contextually, relative to the contents of *b* at the time the claim is made. For this reason, the expert-relative notion of evidence can be useful in studying the history of science, particularly in cases where evidential claims are sound, relative to what was known by the experts at the time they were made, but unsound given what we know today. For example, in the second century A.D. Ptolemy argued that the Earth does not rotate about its axis (*h*). He claimed that his evidence for *h* was that when a stone is tossed up vertically, it is observed to land straight down in the same position from which it was thrown (*e*). Ptolemy asserted that *e* was a reason to believe *h* because, if the Earth were rotating eastward, then the stone should be observed to land west of its starting point (because the Earth would have rotated slightly during the time the stone was in the air). Relative to his knowledge and beliefs *e* increased his subjective degree of belief in *h* (he already had a high prior degree of belief in *h*); further, believing in *h* to the degree he did was coherent with Ptolemy's complete set of beliefs. So, with the subjective Bayesians, we can say that *e* was Ptolemy's *personal* reason for believing *h*. But we can make a stronger evidential claim than this. It is also the case that, relative to the total body of knowledge of his time (which did not include knowledge of inertial motion), *e* was a reason

for *anyone* to believe *h*. Relative to the contents of *b* at time *t*, *e* was an *impersonal* reason to believe *h*. So we can use this expert-relative notion of evidence to evaluate Ptolemy's evidential claim in the impersonal sense, even though, relative to our current knowledge, *e* is not an impersonal reason to believe *h*.²²

This nonobjective concept of evidence, then, is able to capture the impersonal nature of evidence at a given time in science. Yet, like the objective notion of evidence, this expert-relative view is incompatible with the historical thesis. Even on this view, which is relativized to what is known by the experts, *e* can be evidence for *h* regardless of whether any person knows that *e* is true. For example, given that current medical knowledge includes knowledge that spots of a certain kind are always accompanied by measles, the presence of these spots on John's body (*e*) is a reason to believe that John has the measles (*h*). It is evidence for *h* because the fact that he has those spots (*e*), given what is now noncontroversially known about such spots (*b*), gives a reason for believing that *h* is true. This is so even if there is no person, as yet, who knows that John has the spots. Hence, this *e* is evidence even if no person knows that *e* is true. Moreover, *e* can also be evidence on the expert-relative view if *e* is known to be true (either by one person or by "the experts"). For example, prior to Newton's formulation of his law of universal gravitation, it was known that unsupported bodies fall. This information *e* was part of *b* (it was known by the experts); still, this *e* was evidence for Newton's universal theory. On the expert-relative theory of evidence, the time at which *e* is known to be true by anyone or even by everyone is not relevant to its status as evidence for *h*.

■ | Scientific Evidence and the Historical Thesis

I have shown that while the historical thesis is compatible with a personal theory of evidence, such as the subjective Bayesian view, any purely personal theory is seriously flawed in not capturing the nature of evidence as it is used in science. In science, evidence is an impersonal concept. On impersonal concepts of evidence, such as the expert-relative view and the objective views, some *e* can be evidence for *h* whether or not any person knows or believes that *e* is true. Hence the *time* at which *e* is known is not relevant to whether *e* is evidence for *h*, as the historical thesis holds.

Some might argue that *e* cannot be evidence for any hypothesis if no one knows that *e* is true, and might suggest *prima facie* rejection of theories of evidence which entail this conclusion. This response, however, conflates the distinction between *being* evidence and *being used as* evidence. Suppose a doctor wants to consider the hypothesis "this patient has AIDS." The doctor must decide whether or not there is evidence for this

hypothesis; that is, whether there is a reason to believe that the patient has AIDS. The doctor then performs a blood test which discovers the presence of the HIV virus. While it is still not *certain* that the patient does have AIDS (the virus is present before AIDS develops, and in fact not all infected people ever develop AIDS), the doctor clearly does have some reason for believing that the person has AIDS; and this reason exists even before the doctor realizes that it does.* Of course, it is only after the presence of the virus is detected that the doctor *knows* there is evidence, or can *invoke* this fact as evidence, but these are different matters.

This is why in science, as in criminology, we speak of *discovering*—rather than inventing or creating—evidence. When a detective finds a suspect's fingerprints at the crime scene, we say that she has *discovered* evidence (provided, of course, that she has not planted the prints there herself. In this case, we say she has fabricated evidence, not that she has found evidence.) Similarly, scientists discover evidence that existed before it was found. Given the presence of HIV in a patient's blood there is a reason to believe the patient has AIDS, just as the presence of the suspect's fingerprints on the bloody chainsaw near the decapitated body is a reason to believe the suspect committed the murder, even before the virus or the prints are discovered. This is why detectives can speak of destroying evidence by handling objects before they are dusted for fingerprints; that is, destroying the evidence that exists but needs yet to be discovered. On the other hand, scientists can also discover an evidential relation between some information which was previously known, and a hypothesis. For example, doctors discovered that certain kinds of spots, which were already known to occur, were evidence that a patient has measles. If some information gives a reason to believe *h*, then it is evidence for *h* whether it is

* As with any viral infection, we have to distinguish three things: the patient having antibodies to the virus, the presence of the virus in the patient's body, and the patient having symptoms characteristic of the disease (or "syndrome") caused by the virus. Most blood tests for AIDS (the ELISA test, for example) detect antibodies. For several weeks immediately after infection with the human immunodeficiency virus (HIV), the antibody test is likely to be negative even though the virus is present in the patient's body. Also, a patient can test positive and yet be virus-free, either because the test yielded a false positive or, as in the case of some infants born to seropositive women, because the antibodies, but not the virus, are present in the patient's body. In addition, a person can be infected with HIV for ten years or more before manifesting the symptoms of full-blown AIDS. Presumably, Snyder is referring to those rare individuals who, at the time of her article (1994), had been seropositive for HIV for many years but had not developed (and perhaps still have not developed) full-blown AIDS. Such cases are of intense interest to AIDS researchers, since they suggest that some people have something in their bodies that slows down the rate at which the virus replicates. The definition of AIDS used by the Centers for Disease Control in Atlanta, which has changed with advances in our knowledge, now includes HIV infection as a necessary (but not a sufficient) condition.

discovered to be true after *h* is invented or before it is—whether *e* is new evidence or old.²³

I will close by paraphrasing Gertrude Stein: “evidence is evidence is evidence.” The time at which some information is known relative to the forming of a theory is as irrelevant to its evidential value as is the time of day a rose is smelled to its status of being a rose.²⁴

■ | Notes

1. Wesley Salmon (1966) similarly argues that, when a “daring” prediction is successful, “. . . it is not likely to come out right unless we have hit upon the correct hypothesis. Confirming instances are not likely to be forthcoming by sheer chance” (119).

2. John Worrall (1989) uses a version of this “no coincidence” argument in presenting his “heuristic” theory of evidence. A theory’s empirical success, he claims, can be explained either by chance (an option Worrall rejects), by the theory’s truth, or by the theory being “engineered” by the ingenuity of its creator. His heuristic approach is intended to preclude the third option, so that the only plausible option is that the theory is true. My arguments against views requiring either new or old evidence likewise work against Worrall’s theory, insofar as it considers relevant the time at which *e* was known by *h*’s creator relative to *h*’s invention. See also Gardner (1982).

3. See Popper (1959), especially 40–42. In a related vein Giere has more recently argued that “if the known facts were used in constructing the model and were thus built into the resulting hypothesis . . . then the fit between these facts and the hypothesis provides no evidence that the hypothesis is true [since] these facts had no chance of refuting the hypothesis” (1984, 161).

4. Indeed, the more unlikely the prediction is, the more it increases the probability of the hypothesis when it is confirmed; that is, the smaller the figure in the denominator [$p(e)$], the larger $p(h/e)$. Hence, successful predictions of surprising phenomena (what is contrary to expectation) count as the best evidence for a theory, on the positive-relevance view.

5. See Brush (1989), (1990), (1992), and (1993).

6. Brush argues for both a historical and a logical claim. The historical claim is that scientists do, in fact, prefer explanations to predictions as evidence. But he also presents a logical claim aimed against “Popper’s thesis”: that is, Brush argues that scientists *should* prefer explanations, because only explanations can be “reliable” evidence.

7. But note that there are other important reasons why the 1919 eclipse results may not have been “reliable” evidence. For one thing, the values obtained scatter around Einstein’s predicted value, with probable errors large enough to include it; hence it is not obvious that the results *do* confirm the exact prediction. It is telling in this regard that it took four months of calculations before the scientific team making the observations concluded that the results *did* confirm the predic-

tion. The eclipse occurred May 29; Einstein was not notified of the success until September 27; and the results were not announced to the public media until early November.

8. For discussions of "historical" views of confirmation, see Musgrave (1974) and McMullin (1979).

9. I am agreeing with the intuitive notion of evidence put forth by Achinstein (1983), although not with the details of his particular theory.

10. "Reasonableness to believe" is generally, and I think correctly, linked to the probability or the likelihood of the hypothesis. If e makes it reasonable to believe h , then h is very probable, given e . Proponents of the positive-relevance view discussed in the first section argue, by contrast, that e must increase the probability of h , and does not need to make h very probable.

11. *Science*, 260 (28 May, 1993), 1255–6, 1273–8.

12. To deny this role of evidence, and claim that scientists cannot use evidence as a means of deciding among theories, would be to claim that science is nonrational; that is, that scientists must always choose which theories to believe on the basis of social, political, and/or rhetorical reasons which have nothing to do with the attempt to discover how the world really works. While it may be true that scientists sometimes do, in fact, choose among theories in this nonrational way, it surely goes too far to claim that they do and should *only* choose which theories to believe on the basis of nonrational considerations.

13. See the following articles collected in Achinstein (1983): Carl G. Hempel, "Studies in the Logic of Confirmation," 10–43; Rudolf Carnap, "The Concept of Confirming Evidence," 79–94; Clark Glymour, "Relevant Evidence," 124–44; and Peter Achinstein, "Concepts of Evidence," 145–74.

14. Note that these are Achinstein's conditions for *potential* evidence.

15. For a full discussion of the subjective Bayesian position, see Howson and Urbach (1989).

16. See Glymour (1980), 85–93.

17. However, Howson and Urbach (1989) have argued that the predictivist position is not entailed by their view, because the conditional probability $p(e/h)$ should be set for the counterfactual situation where one does not know, in advance, that e is true. Using this value of $p(e/h)$ to determine $p(h/e)$, the evidential value of e for h measures the extent to which e *would* increase a person's degree of belief. So on their construal of the subjective Bayesian position, the historical thesis is rejected (see 270–75).

18. The Bayesians claim that, in fact, people do come to have the same or similar degrees of belief in certain hypotheses; as more confirming instances of a universal law are observed, it becomes more likely that we will agree in expecting the next instance to also be a confirming one (there will be a "convergence of belief"). But this probabilistic claim has been defeated by John Earman (see his 1985). Moreover, our shared belief would extend only to the next instance, not to the belief in the truth of the hypothesis itself.

19. See *Principia*, Book III. For a more lengthy discussion of Newton and "the phenomena," see Achinstein (1991), 33–5.

20. To flesh out this view further, it would be necessary to explain how exactly the experts are to be distinguished from the nonexpert nonbelievers. That is, we need a way to determine when disagreement signals ongoing controversy among scientific experts (as in the AIDS research community), and when it signals merely the disagreement of the nonexpert (as when the Church continued to deny the findings of Galileo, even after scientific controversy ended). This could be done to some extent by examining publications in respected scientific journals, membership in certain associations, etc.

21. This view is not to be confused with the standard way of relativizing evidential claims to "background information." When objective theories relativize claims to background information, as Carnap, Achinstein, and others do, this information *b* is taken to consist in any proposition or set of propositions at all—it is not limited to what is known by anyone, as I am suggesting here. When the subjective Bayesian view is relativized to background information, *b* consists in the knowledge and beliefs of a particular person; this also differs from relativization to a "total body of knowledge" of a community of experts.

22. The observed straight path of the landing stone is ~~due~~ not to the earth's immobility, but (as we now know) to inertia—to the tendency of a body to persist in a motion it has acquired. The stone continues to move eastward with the earth even when it is tossed up a few feet.

23. As the decapitated head example suggests, some *e* can be evidence for *h* even if it is neither explained nor entailed by *h*. That is, evidence does not *need* to consist in either an explanation or a prediction. From the hypothesis that John is dead, it does not follow as a predicted consequence that John's body is in two pieces. And the hypothesis that John is dead does not explain the fact that his head is no longer attached to his body (although it is certainly consistent with this fact). Yet the fact that his body is in this state is about as good a reason as we could have to believe that John is dead.

24. I am grateful to Peter Achinstein for many helpful suggestions. I would also like to thank Robert Rynasiewicz for comments on an earlier draft of this paper.

■ | References

Achinstein, P. (ed.) 1983. *The Concept of Evidence*. New York: Oxford University Press.

———. 1991. *Particles and Waves*. New York: Oxford University Press.

Brush, S. G. 1989. "Prediction and Theory Evaluation: The Case of Light Bending," *Science* 246: 1124–29.

———. 1990. "Prediction and Theory Evaluation: Alfven on Space Plasma Phenomena," *Eos* 71 (2): 19–33.

- . 1992. "How Cosmology Became a Science," *Scientific American* 267 (2): 62–70.
- . 1993. "Prediction and Theory Evaluation: Cosmic Microwaves and the Revival of the Big Bang," *Perspectives on Science* 1: 565–602.
- Earman, J. 1985. "Concepts of Projectibility and Problems of Induction," *Nous* 19: 521–35.
- Gardner, M. R. 1982. "Predicting Novel Facts," *British Journal for the Philosophy of Science* 33: 1–15.
- Giere, R. N. 1984. *Understanding Scientific Reasoning*, 2nd edition. New York: Holt, Rinehart and Winston.
- Glymour, C. 1980. *Theory and Evidence*. Princeton: Princeton University Press.
- Gooding, D. et al. (eds.) 1989. *The Uses of Experiment*. Cambridge: Cambridge University Press.
- Harman, G. 1965. "The Inference to the Best Explanation," *Philosophical Review* 74 (1): 88–95.
- Howson, C. and P. Urbach. 1989. *Scientific Reasoning: The Bayesian Approach*. LaSalle, IL: Open Court Press.
- McMullin, E. 1979. "The Ambiguity of Historicism," *Current Research in Philosophy of Science*: 55–83.
- Musgrave, A. 1974. "Logical versus Historical Theories of Confirmation," *British Journal for the Philosophy of Science* 25: 1–23.
- Newton, I. 1966. *Mathematical Principles of Natural Philosophy* (the *Principia*), 2 vols., trans A. Motte, rev. F. Cajori. Berkeley and Los Angeles: University of California Press.
- Popper, K. R. 1959. *The Logic of Scientific Discovery*. London: Hutchinson and Co.
- Salmon, W. 1966. *The Foundations of Scientific Inference*. Pittsburgh: University of Pittsburgh Press.
- Whewell, W. 1857. *History of the Inductive Sciences*, 3rd edition, in three volumes. London.
- . 1860. *On the Philosophy of Discovery*. London.
- Worrall, J. 1989. "Fresnel, Poisson and the White Spot: The Role of Successful Predictions in the Acceptance of Scientific Theories," in Gooding et al.

Explanation v. Prediction: Which Carries More Weight?

1 | The Historical Thesis of Evidence

According to a standard view, predictions of new phenomena provide stronger evidence for a theory than explanations of old ones. More guardedly, a theory that predicts phenomena that did not prompt the initial formulation of that theory is better supported by those phenomena than is a theory by known phenomena that generated the theory in the first place. So say various philosophers of science, including William Whewell (1847) in the 19th century and Karl Popper (1959) in the 20th, to mention just two.

Stephen Brush takes issue with this on historical grounds. In a series of fascinating papers he argues that generally speaking scientists do not regard the fact that a theory predicts new phenomena, even ones of a kind totally different from those that prompted the theory in the first place, as providing better evidential support for that theory than is provided by already known facts explained by the theory. By contrast, Brush claims, there are cases, including general relativity and the periodic law of elements, in which scientists tend to consider known phenomena explained by a theory as constituting much stronger support than novel predictions.¹

Both the predictionist and the explanationist are committed to an interesting historical thesis about evidence, viz.

Historical thesis: Whether some claim *e*, if true, is evidence for an hypothesis *h*, or how strong that evidence is, depends on certain historical facts about *e*, *h*, or their relationship.

For example, whether, or the extent to which, *e* counts as evidence for *h*

FROM D. Hull, M. Forbes, and R. M. Burian, eds., *PSA 1994*, vol. 2 (East Lansing, Mich.: Philosophy of Science Association, 1994), 156–64.

depends on whether *e* was known before or after *h* was formulated. Various historical positions are possible, as Alan Musgrave (1974) noted years ago in a very interesting article. On a simple predictionist view (which Musgrave classifies as “purely temporal”) *e* supports *h* only if *e* was not known when *h* was first proposed. On another view (which Musgrave attributes to Zahar (1973) and calls “heuristic”), *e* is evidence for *h* only if when *h* was first formulated it was not devised in order to explain *e*. On yet a third historical view (which Musgrave himself accepts), *e* is evidence for some theory *T* only if *e* cannot be explained by a “predecessor” theory, i.e., by a competing theory which was devised by scientists prior to the formulation of *T*. These views, and other variations, are all committed to the historical thesis.

Is the historical thesis true or false? I propose to argue that it is sometimes true, and sometimes false, depending on the type of evidence in question. Then I will consider what implications, if any, this has for the debate between Brush and the predictionists.

Before beginning, however, let me mention a curious but interesting fact about various well-known philosophical theories or definitions of evidence. As Laura Snyder (1994) points out in a perceptive paper entitled “Is Evidence Historical?”, most such theories, including Carnap’s (1962) *a priori* theory of confirmation, Hempel’s (1945) satisfaction theory, Glymour’s (1980) bootstrap account, and the usual hypothetico-deductive account, are incompatible with the historical thesis. They hold that whether, or the extent to which, *e* is evidence for, or confirms, *h* is an objective fact about *e*, *h*, and their relationship. It is in no way affected by the time at which *h* was first proposed, or *e* was first known, or by the intentions with which *h* was formulated. Defenders of these views must reject both the predictionist and the explanationist claims about evidence. They must say that whether, or the extent to which, *e* supports *h* has nothing to do with whether *e* was first formulated as a novel prediction from *h* or whether *e* was known before *h* and *h* was constructed to explain it.

Accordingly, we have two extreme or absolutist positions. There is the position, reflected in the historical thesis, that evidence is always historical (in the sense indicated). And there is a contrasting position, reflected in certain standard views, that evidence is never historical. Does the truth lie at either extreme? Or is it somewhere in the middle?

2 | Selection Procedures

Suppose that an investigator decides to test the efficacy of a certain drug *D* in relieving symptoms *S*. The hypothesis under consideration is

h: Drug D relieves symptoms S in approximately 95% of the cases.

The investigator may test drug D by giving it to persons suffering from S and by giving a placebo to other persons suffering from S (the "control group"). In deciding how to proceed, the investigator employs what I will call a "selection procedure," or rule, determining how to test, or obtain evidence for, an hypothesis, in this case determining which persons he will select for his studies and how he will study them.

For example, here is one of many possible selection procedures (SP) for testing **h**:

SP 1: Choose a sample of 2000 persons of different ages, sexes, races, and geographical locations, all of whom have symptoms S in varying degrees; divide them arbitrarily into 2 groups; give one group drug D and the other a placebo; determine how many in each group have their symptoms relieved.

Now, suppose that a particular investigator uses this (or some other) selection procedure and obtains the following result:

e: In a group of 1000 persons with symptoms S taking drug D, 950 persons had relief of S; in a control group of 1000 S-sufferers not taking D but a placebo none had symptoms S relieved.

The first thing to note about this example is that whether the report **e** supports hypothesis **h**, or the extent to which it does, depends crucially on what selection procedure was in fact used in obtaining **e**. Suppose that instead of SP1 the following selection procedure had been employed:

SP 2: Choose a sample of 2000 females aged 5 all of whom have symptoms S in a very mild form; proceed as in SP1.

If result **e** had been obtained by following SP2, then **e**, although true, would not be particularly good evidence for **h**, certainly not as strong as that obtained by following SP1. The reason, of course, is that SP1, by contrast with SP2, gives a sample that is varied with respect to two factors that may well be relevant: age of patient and severity of symptoms. (Hypothesis **h** does not restrict itself to 5 year old girls with mild symptoms, but asserts a cure-rate for the general population of sufferers with varying degrees of the symptoms in question.)

This means that if the result as described in **e** is obtained, then whether, or to what extent, that result confirms the hypothesis **h** depends crucially on what selection procedure was in fact used in obtaining **e**. That is, it depends on an historical fact about **e**: on how in fact **e** was

obtained. If *e* resulted from following SP1, then *e* is pretty strong evidence for *h*; if *e* was obtained by following SP2, then *e* is pretty weak evidence for *h*, if it confirms it at all. Just by looking at *e* and *h*, and even by ascertaining that *e* is true, we are unable to determine to what extent, if any, *e* supports *h*. We need to invoke "history."

To nail down this point completely, consider a third selection procedure:

SP 3: Choose a sample of 2000 persons all of whom have *S* in varying degrees; divide them arbitrarily into 2 groups; give one group drugs *D* and *D'* (where *D'* relieves symptoms *S* in 95% of the cases and blocks possible curative effects of *D* when taken together); give the other group a placebo.

Consider once more result *e* (which, again, let us suppose, obtains). In this case *e* supports *h* not at all. And, again, whether this is so cannot be ascertained simply by examining the propositions *e*, *h*, or their "logical" relationship. We need to know an historical fact about *e*, viz. that the information it (truly) reports was obtained by following SP3.

So far then we seem to have support for the historical thesis about evidence. Can we generalize from examples like this to all cases? Can we say that for any true report *e*, and any hypothesis *h*, whether, or to what extent, *e* is evidence for *h* depends upon historical facts about how *e* was obtained? No, we cannot.

Consider another very simple case. Let *e* be the following report, which is true:

e = In last week's lottery, 1000 tickets were sold, of which John owned 999 at the time of the selection of the winner; this was a fair lottery in which one ticket was selected at random.

h = John won the lottery.

In an attempt to obtain information such as *e* to support *h* different rules or "selection procedures" might have been followed, e.g.,

SP 4: Determine who bought tickets, and how many, by asking lottery officials.

SP 5: Determine this by standing next to the person selling tickets.

SP 6: Determine this by consulting the local newspaper, which publishes this information as a service to its readers.

Let us suppose that following any of these selection procedures results in a true report *e*. (And, as in the symptoms case, we may suppose that following any of these procedures is a reasonable way to establish whether *e* is true.) But in this case, unlike the drug example, which selection procedure was in fact followed is completely irrelevant in determining whether, or to what extent, *e* is evidence for *h*. In this case, unlike the drug example, we do not need to know how information *e* was obtained to know that *e* (assuming it is true) is very strong evidence for *h*. Nor do we need to know any other historical facts about *e*, *h*, or their relationship. (In particular, contrary to both the predictionist and explanationist views, we do not need to know when *e* was first known relative to when *h* was first formulated; i.e., we do not need to know whether *e* was explained or predicted. But more of this later when these two historical views are examined more fully.) Accordingly, we have a case that violates the historical thesis of evidence.

Since examples similar to each of the two above can be readily constructed, we may conclude that there are many cases that satisfy the historical thesis of evidence, and many others that fail to satisfy it. Is there a general rule for deciding which do and which do not?

Perhaps our two examples will help generate such a rule. In the drug case the evidence report *e* is historical in an obvious sense: it reports the results of a particular study made at some particular time and place. But this is clearly not sufficient to distinguish the cases, since the evidence report *e* in the lottery case is also historical: it reports facts about a particular lottery, who bought tickets, and when. So, I submit, what distinguishes the cases is not the historical character of the evidence, but something else.

I shall say that a putative evidence statement *e* is *empirically complete* with respect to an hypothesis *h* if whether, or to what extent, *e* is evidence for, or confirms, *h* depends just on what *e* reports, what *h* says, and the relationship between them. It does not depend on any additional empirical facts—e.g., facts about when *e* or *h* were formulated, or with what intentions, or on any (other) facts about the world. In the drug example, *e* is not empirically complete with respect to *h*: whether, or to what extent, *e* supports *h* depends on how the sample reported in *e* was selected—empirical information not contained in *e* or *h*.² By contrast, in the lottery example, *e* is empirically complete with respect to *h*: whether, and to what extent, *e* supports *h* in this case does not depend on empirical facts in addition to *e*. To determine whether, and how much, *e* supports *h* in this case we do not need any further empirical investigation. To be sure, additional empirical inquiry may unearth new information *e'* which is such that both *e* and *e'* together do not support *h* to the same extent that *e* by itself does. But that is different. In the drug but not the lottery example information in addition to *e* is necessary to determine the extent to which *e* itself supports *h*. In the drug case we cannot legitimately say whether or

to what extent the report *e* supports the efficacy of drug *D* unless we know how the patients described in *e* were selected. In the lottery case information about how purported evidence was obtained is irrelevant for the question of whether or how strongly that evidence, assuming its truth, supports *h*.

So we have one important difference between the two examples. Is this enough to draw a distinction between cases that satisfy the historical thesis of evidence and those that do not? Perhaps not. There may be cases in which *e* is empirically incomplete with respect to *h*, but in which empirical facts needed to complete it are not historical. Consider

e = Male crows are black.

h = Female crows are black.

One might claim that whether, or the extent to which, *e* supports *h* in this case depends on empirical facts in addition to *e*. If, e.g., other species of birds generally have different colors for different sexes, then *e* does not support *h* very much. If other species generally have the same color for both sexes, then *e* supports *h* considerably more. But these additional facts are not "historical," at least not in the clear ways of previous examples. (I construe "other species of birds generally have different colors for different sexes" to be making a general statement, and not to be referring to any particular historical period.) If this is granted, then we need to add a proviso to the completeness idea above.

There are cases (including our drug example) in which a putative evidence claim *e* is empirically incomplete with respect to an hypothesis *h*, where determining whether, or to what extent, *e* supports *h* requires determining the truth of some historical fact. I shall speak of these as historical evidence cases. They satisfy the historical thesis of evidence. By contrast, there are cases in which a putative evidence claim *e* is empirically complete with respect to hypothesis *h* (e.g., our lottery case); and there may be cases in which a putative evidence claim *e*, although empirically incomplete with respect to *h*, can be settled without appeal to historical facts (possibly the crow example). Cases of the latter two sorts violate the historical thesis of evidence. What implications, if any, does this hold for whether predictions or explanations provide better confirmation?

3 | Predictions v. Explanations

Let us return to the original question proposed by Brush. Do predictions of novel facts provide stronger evidence than explanations of old ones, as Whewell and Popper claim? Or is the reverse true? My answer is this: Sometimes a prediction provides better evidence for an hypothesis, some-

times an explanation does, and sometimes they are equally good. Which obtains has nothing to do with the fact that it is a prediction of novel facts or that it is an explanation of known ones.

To show this, let us begin with a case that violates the historical thesis of evidence. Here it should be easy to show that whether the putative evidence is known before or after the hypothesis is formulated is irrelevant for confirmation. Let the hypothesis be

h = This coin is fair, i.e., if tossed in random ways under normal conditions it will land on heads approximately half the time in the long run.

e = This coin is physically symmetrical, and in a series of 1000 random tosses under normal conditions it landed on heads approximately 500 times.

We might reasonably take e to be empirically complete with respect to h . Accordingly, whether e supports h , and the extent to which it does, does not depend on empirical facts other than e . In particular, it does not depend on when, how, or even whether e comes to be known, or on whether e was known first and h then formulated, or on whether h was conceived first and e then stated as a prediction from it. Putative evidence e supports hypothesis h and does so (equally well) whether or not e is known before or after h was initially formulated, indeed whether or not e is ever known to be true.

So let us focus instead on cases that satisfy the historical thesis of evidence. We might suppose that at least in such cases explanations (or predictions) are always better for confirmation. Return once again to our drug hypothesis:

h = Drug D relieves symptoms S in approximately 95% of the cases.

Consider now two evidence claims, the first a prediction about an unknown future event, the second a report about something already known:

e_1 = In the next clinical trial of 1000 patients who suffer from symptoms S and who take D approximately 950 will get some relief.

e_2 = In a trial that has already taken place involving 1000 patients with S who took D (we know that) approximately 950 got some relief.

On the prediction view, e_1 is stronger evidence for h than is e_2 . On the explanation view it is the reverse. And to sharpen the cases let us suppose that e_2 , by contrast to e_1 , was not only known to be true prior to the

formulation of h , but that h was formulated with the intention of explaining e_2 . Which view is correct? Neither one.

Let us take the prediction case e_1 first. Whether, and to what extent, e_1 (if true) supports h depends on empirical facts in addition to e_1 . In this case it depends on the selection procedure to be used in the next clinical trial. Suppose this selection procedure calls for choosing just 5 year old girls with very mild symptoms who in addition to D are also taking drug D' which ameliorates symptoms S in 95% of the cases and potentially blocks D from doing so. Then e_1 would be very weak evidence for h , if it supports it at all. This is so despite the fact that e_1 is a correct prediction from h , one not used in generating h in the first place. By contrast, suppose that the selection procedure used in the past trial mentioned in e_2 is much better with respect to h . For example, it calls for choosing humans of both sexes, of different ages, with symptoms of varying degrees, who are not also taking drug D' . Then e_2 would be quite strong evidence for h , much stronger than what is supplied by e_1 . In such a case, a known fact explained by h would provide more support for h than a newly predicted fact would.

Obviously the situations here can be reversed. We might suppose that the selection procedure used to generate the prediction of e_1 is the one cited in the previous paragraph as being used to generate e_2 (and vice versa). In this situation a newly predicted fact would provide more support for h than an already explained one.

In these cases what makes putative evidence have the strength it does has nothing to do with whether it is being explained or predicted. It has to do with the selection procedure used to generate that evidence.¹ In one situation—whether it involves something that is explained or predicted—we have a putative evidence statement generated by a selection procedure that is a good one relative to h ; in the other case we have a flawed selection procedure. This is what matters for confirmation—not whether the putative evidence is being explained or predicted.

4 | Brush Redux

Brush is clearly denying a general predictionist thesis. By contrast he cites cases in which scientists themselves regarded known evidence explained by a theory as stronger support for that theory than new evidence that was successfully predicted. And he seems to imply that this was reasonable. He offers an explanation for this claim, viz. that with explanations of the known phenomena, by contrast with successful predictions of the new ones, scientists had time to consider alternative theories that would generate these phenomena. Now, even if Brush does not do so, I want to

extend this idea and consider a more general explanationist view that is committed to the following three theses that Brush invokes for some cases:

- 1 A selection procedure for testing a hypothesis *h* is flawed, or at least inferior to another, other things equal, if it fails to call for explicit consideration of competitors to *h*.
- 2 The longer time scientists have to consider whether there are plausible competitors to *h* the more likely they are to find some if they exist.
- 3 With putative evidence already known before the formulation of *h* scientists have (had) more time to consider whether there are plausible competitors to *h* than is the case with novel predictions.

I would challenge at least the first and third theses. In my first example, selection procedure 1 for the drug hypothesis does not call for explicitly considering competitors to that hypothesis. Yet it does not seem flawed on that account, or inferior to one that does. However, even supposing it were inferior, whether or not a selection procedure calls for a consideration of competitors is completely irrelevant to whether the putative evidence claim is a prediction or a known fact being explained. In the case of a prediction, no less than that of an explanation, the selection procedure may call for a consideration of competitors.

For example, in our drug case, where *h* is "Drug D relieves symptoms S in approximately 95% of the cases," and *e* is the prediction "In the next clinical trial of 1000 patients suffering from symptoms S who take D, approximately 950 will get some relief," the selection procedure to be used for the next clinical trial might include the rule

In conducting this next trial, determine whether the patients are also taking some other drug which relieves S in approximately 95% of the cases and which blocks any effectiveness D might have.

Such a selection procedure calls for the explicit consideration of a competitor to explain *e*, viz. that it will be some other drug, not D, that will relieve symptoms S in the next trial. This is so even though *e* is a prediction. Moreover, to respond to the third thesis about time for considering competitors, an investigator planning a future trial can have as much time as she likes to develop a selection procedure calling for a consideration of a competing hypothesis. More generally, in designing a novel experiment to test some hypothesis *h* as much time may be spent in precluding competing hypotheses that will explain the test results as is spent in considering competing hypotheses for old data.

5 | Thomson v. Hertz

Finally, let me invoke an example more recognizably scientific. It involves a dispute between Heinrich Hertz and J. J. Thomson over the nature of cathode rays.⁴ In experiments conducted in 1883 Hertz observed that the cathode rays in his experiments were not deflected by an electrical field. He took this to be strong evidence that cathode rays are not charged particles (as the English physicist William Crookes had concluded), but some type of ether waves. In 1897 J. J. Thomson repeated Hertz's experiments but with a much higher evacuation of gas in the cathode tube than Hertz had been able to obtain. Thomson believed that when cathode rays pass through a gas they make it a conductor, which screens off the electric force from the charged particles comprising the cathode rays.⁵ This screening off effect will be reduced if the gas in the tube is more thoroughly evacuated. In Thomson's 1897 experiments electrical deflection of the cathode rays was detected, which Thomson took to be strong evidence that cathode rays are charged particles.

Here, however, I want to consider the evidential report of Hertz in 1883, not of Thomson in 1897. Let

e = In Hertz's cathode ray experiments of 1883 no electrical deflection of cathode rays was detected.

h = Cathode rays are not electrically charged.

Hertz took e to be strong evidence for h. In 1897 Thomson claimed, in effect, that Hertz's results as reported in e did not provide strong evidence for h, since Hertz's experimental set-up was flawed: He was employing insufficiently evacuated tubes. To use my previous terminology, Thomson was claiming that Hertz's selection procedure for testing h was inadequate.⁶

Here we can pick up on a point emphasized by Brush. Hertz, we might say, failed to use a selection procedure calling for considering a competitor to h to explain his results (viz. that cathode rays are charged particles, but that the tubes Hertz was using were not sufficiently evacuated to allow an electrical force to act on these particles). But—and this is the point I want to emphasize—in determining whether, or to what extent, Hertz's putative evidence e supports his hypothesis h, it seems to be irrelevant whether Hertz's e was a novel prediction from an already formulated hypothesis h or an already known fact to be explained by h. Hertz writes that in performing the relevant experiments he was trying to answer two questions:

Firstly: Do the cathode rays give rise to electrostatic forces in their neighbourhood? Secondly: In their course are they affected by external electrostatic forces? (Hertz 1896, p. 249)

In his paper he did not predict what his experiments would show. Nor were the results of his experiments treated by him as facts known before he had formulated his hypothesis *h*. Once he obtained his experimental result he then claimed that they supported his theory:

As far as the accuracy of the experiment allows, we can conclude with certainty that no electrostatic effect due to the cathode rays can be perceived. (p. 251)

To be sure, we might say that Hertz's *theory* itself predicted some such results, even if Hertz himself did not (i.e., even if Hertz did not himself draw this conclusion before getting his experimental results). But even if we speak this way, Hertz did not claim or imply that his experimental results provide better (or weaker) support for his theory because the theory predicted them before they were obtained. Nor did Thomson in his criticism of Hertz allude to one or the other possibility. Whichever it was—whether a prediction or an explanation or neither—Hertz (Thomson was claiming) should have used a better selection procedure. This is what is criticizable in Hertz, not whether he was predicting a novel fact or explaining a known one.

I end with a quote from John Maynard Keynes (1921, p. 305), whose book on probability contains lots of insights. Here is one:

The peculiar virtue of prediction or predesignation is altogether imaginary. The number of instances examined and the analogy between them are the essential points, and the question as to whether a particular hypothesis happens to be propounded before or after their examination is quite irrelevant.

■ | Notes

1. To what extent Brush wants to generalize this explanationist position is a question I leave for him to answer. There are passages in his writings that strongly suggest a more general position. For example: "There is even some reason to suspect that a successful explanation of a fact that other theories have already failed to explain satisfactorily (for example, the Mercury perihelion) is more convincing than the prediction of a new fact, at least until the competing theories have had their chance (and failed) to explain it" (1989, p. 1127). In what follows I consider a generalized explanationist thesis.

2. For Carnap (1962) and others, every *e* is empirically complete with respect to every *h*. For these writers, whether, and the extent to which, *c* confirms *h* is an *a priori* matter.
3. Cf. Mayo (1991).
4. See Achinstein (1991), Essays 10 and 11; also Buchwald (1994), ch. 10.
5. See Thomson (1897), p. 107.
6. Lord Rayleigh (1942, pp. 78–9), in a biography of Thomson, made the same claim: "He [Hertz] failed to observe this [electrical] effect, but the design of his experiment was open to certain objections which were removed in a later investigation by Perrin in 1895, directed to the same question. Perrin got definite evidence that the rays carried a negative charge. J. J. Thomson, in a modification of Perrin's experiment showed that if the Faraday cylinder was put out of the line of fire of the cathode, it acquired a charge when, and only when, the cathode rays were so deflected by a magnet as to enter the cylinder." [Note Rayleigh's claim that Perrin (and Thomson) got "definite evidence" that cathode rays carry a negative charge, whereas, by implication, Hertz's experiments did not give "definite evidence" concerning the question of charge.]

Acknowledgement: I am indebted to Laura J. Snyder for very helpful discussions, and to Robert Rynasiewicz for trying to convince me of the error of my ways.

■ | References

- Achinstein, P. (1991), *Particles and Waves*. New York: Oxford University Press.
- Brush, S. (1989), "Prediction and Theory Evaluation: The Case of Light Bending," *Science* 246: 1124–29.
- Buchwald, J. (1994), *The Creation of Scientific Effects*. Chicago: University of Chicago Press.
- Carnap, R. (1962), *Logical Foundations of Probability*. Chicago: University of Chicago Press, 2nd ed.
- Glymour, C. (1980), *Theory and Evidence*. Princeton: Princeton University Press.
- Hempel, C. (1945), "Studies in the Logic of Confirmation," *Mind* 54: 1–26, 97–121.
- Hertz, H. (1896), *Miscellaneous Papers*. London: Macmillan.
- Keynes, J.M. (1921), *A Treatise on Probability*. London: Macmillan.
- Mayo, D. (1991), "Novel Evidence and Severe Tests," *Philosophy of Science* 58: 523–52.
- Musgrave, A. (1974), "Logical versus Historical Theories of Confirmation," *British Journal for the Philosophy of Science* 25: 1–23.
- Popper, K. (1959), *The Logic of Scientific Discovery*. London: Hutchinson.

Rayleigh, R. (1942), *The Life of Sir J. J. Thomson*. Cambridge: Cambridge University Press.

Snyder, L. (1994), "Is Evidence Historical?" *Scientific Methods: Conceptual and Historical Problems*, P. Achinstein and L. Snyder, eds. Malabar, Florida: Krieger.

Thomson, J. J. (1897), "Cathode Rays," *The Electrician* 39: 104–108.

Whewell, W. (1847), *The Philosophy of the Inductive Sciences*. New York: Johnson Reprint, 1967.

Zahar, E. (1973), "Why Did Einstein's Programme Supersede Lorentz's?" *British Journal for the Philosophy of Science* 24: 95–123, 223–62.

- 4.1 Lipton on the Problem of Induction** 495
Underdetermination 496
The Problem of Justification 497
The Problem of Description 501
- 4.2 Popper's Rejection of Induction** 505
Why Induction Cannot Be Justified 505
Why There Is No Logic of Discovery 507
Falsificationism and Corroboration 507
- 4.3 Salmon's Criticism of Popper's Anti-Inductivism** 508
The Problem of Rational Prediction 509
The Inadequacy of Popperian Corroboration 510
- 4.4 Hempel on Criteria of Confirmation** 510
Quantity, Variety, and Precision of Evidence 511
Novel Prediction 512
Theoretical Support 516
Simplicity 518
The Probability of Hypotheses 523
- 4.5 Snyder on Evidence, Explanation, and Novel Prediction** 525
The Historical Thesis of Evidence 525
Three Arguments for Predictionism 525
Two Arguments for Explanationism 528
Why the Historical Thesis Is False 531
- 4.6 Achinstein's Rejection of Predictionism and Explanationism** 533
Selection Procedures 534
Prediction versus Explanation 535
Brush Redux 536
Thomson versus Hertz 537
- 4.7 Summary** 538

4 | COMMENTARY

4.1 | Lipton on the Problem of Induction

One important class of inductive arguments consists of inferences from a sample to a larger population from which the sample has been selected. For example, a biologist observes a number of reptiles (e.g., garter snakes, skinks, and tortoises) and observes (via dissection) that each of these creatures has a three-chambered heart. From this she infers that all reptiles (including the many thousands that have not been and may never be observed) also have a three-chambered heart. Alternatively, the biologist might study several hundred lizards, find that 98 percent of them are carnivorous, and conclude that 98 percent of lizards are carnivorous. These arguments are typical examples of *inductive generalization*. When we verify some of the consequences of a scientific theory and infer that the theory is (probably) true, our inference from those verified consequences to the truth of the theory is like arguing from a sample—those verified consequences of the theory—to a population—the entire set of the theory's consequences. In concluding that all the theory's consequences are true we are, in effect, concluding that the theory is true.

Inductive arguments come in all shapes and sizes, inductive generalization being but one among them. Other types include *induction to a particular*, *statistical syllogism*, and *arguments from analogy*. Inductions to a particular conclude that the next observed A will be B from the premise that all observed As have been B. In a statistical syllogism, the same conclusion is drawn from the premise that a high percentage of As have been B. Arguments from analogy infer that an individual with properties R, S, and T will also have property U, from the fact that several other individuals with properties R, S, and T have been found to have property U.

Lipton identifies two problems concerning induction: the problems of description and justification. The problem of description is the problem of identifying the general principles we follow when we make inductive inferences. The problem of justification is the problem of explaining why those general principles are reliable. Typically, general principles are judged reliable if arguments with true premises following those principles will lead to true conclusions most of the time. Lipton argues that both problems arise because in every inductive argument, the conclusion is *underdetermined* by the information in the premises. By their very nature, inductive arguments are ampliative: their conclusions go beyond anything that can be inferred from their premises using deductive logic.

UNDERDETERMINATION

The justification problem—often just called the problem of induction in the philosophical literature—arises from the fact that inductive arguments are deductively underdetermined. Since the premises of inductive arguments do not entail their conclusions, no conclusion of an inductive argument is guaranteed to be true, even if all the premises of the argument are true and known to be true with absolute certainty. What, then, justifies our inferring the truth (or probable truth) of the conclusion of an inductive argument? As Lipton explains in the second section of his article, many philosophers (especially David Hume and Karl Popper) have judged that the correct answer to this question is, “Nothing.” In other words, these philosophers are skeptics about the possibility of justifying induction.

Lipton's second problem, the problem of description, also arises from underdetermination. The problem of description, recall, is to say how inductive inferences in fact proceed—to describe the principles deployed in inductive reasoning, quite independently of whether these principles are reliable. For any group of reasoners and learners, a full description of the evidence and other known principles or rules available to them will often leave open what sort of inferences they in fact draw. Lipton gives two illustrations: Noam Chomsky's theory of how children acquire language and Thomas Kuhn's appeal to shared exemplars to explain how scientists reach agreement about scientific problems. In either case, the inferences made cannot be reconstructed adequately using only the principles of deductive logic and the simple, common forms of inductive argument (such as inductive generalization and arguments from analogy). Thus, in the case of language acquisition, Chomsky concludes that there must be additional principles and rules that children follow. Children are not consciously aware of these rules but they are implicit in their ability to master their native tongue rapidly and accurately. Children acquire the ability to construct indefinitely many new grammatical sentences on the basis of the sample of sentences they hear adults utter. Remarkably, not only are the sentences that form the “premises” of the child's inference relatively few but many of them are also ungrammatical. Nonetheless, young children acquire their first language amazingly quickly. Since any normal child can learn any human language, the implicit rules and principles must be common to all human languages (and thus Chomsky calls them a *universal grammar*) and the principles of this universal grammar must be innate.

Lipton says of both Chomsky and Kuhn that they are “arguing for unacknowledged principles of induction” (413). But the connection between Chomsky's theory of linguistic competence and the problem of description for induction is somewhat tenuous. Acquiring competence in a human language is unlike an inference in important ways. In an inference, one moves from the presumed truth of the premises to the presumed

truth of the conclusion. What premises are presumed to be true in the case of language acquisition? The child is not arguing from the truth of the sentences she hears to the truth of the new sentences she constructs. Even if we imagine the child saying to herself (so to speak) something like "Many, but probably not all, the sentences I have heard are grammatically well formed," it is still unclear what conclusion about a particular language would follow from this premise, for Chomsky's point is that there are too many grammars consistent with the sentence inputs for inductive principles to be of any help. Indeed, the innate grammatical principles Chomsky posits are viewed as *imposing* generalizations on the data (sentence inputs), not as extracting grammatical generalizations from them (as talk of "inductive principles" of inference would suggest). On the face of it, it is unclear that Chomsky's fascinating and influential theory offers a cogent illustration of the problem of description for inductive inferences (even if it does illustrate how underdetermination gives rise to the problem of describing hidden principles generally).

A similar difficulty arises with respect to Lipton's appeal to Kuhn's views on scientific judgment as an illustration of the description problem. For as Kuhn is at pains to point out, he thinks that exemplars (shared examples) are necessary for producing consensus among scientists precisely because no set of rules could guide their reasoning. Consequently, it sounds odd for Lipton to characterize Kuhn as "arguing for unacknowledged principles of induction." (See chapter 2 for a detailed examination of Kuhn's views about scientific reasoning and judgment.)

THE PROBLEM OF JUSTIFICATION

In an essay on Francis Bacon, C. D. Broad described induction as "the glory of Science" but "the scandal of Philosophy."¹ Induction is scandalous because philosophers still cannot agree on the correct response to David Hume's skeptical conclusion that inductive reasoning lacks any kind of epistemic justification. Hume argued that any attempt to show that inductive inferences are (in Lipton's phrase) *truth-tropic* is doomed to fail. As far as Hume was concerned, our addiction to inductive reasoning is merely a blind, nonrational, animal faith: it can be explained psychologically, but if Hume is right, it cannot be justified philosophically.

Lipton describes the justification problem—the problem of answering Hume's skepticism about induction—as arising from underdetermination and circularity. We can illustrate this by considering one type of inductive inference, induction to a particular (IP):

All observed As have been B.

The next individual is an A.

The next individual will also be a B.

Now suppose that we attempt to show that IP is a justified form of inference. Hume's all-purpose argument that this is impossible runs as follows:

- 1 If IP can be shown to be justified, then there is an argument that shows it.
 - 2 Arguments are either deductively valid or inductive.
 - 3 No deductively valid argument can justify IP (because of underdetermination)
 - 4 No inductive argument can justify IP (because of circularity).
-
- 5 IP cannot be shown to be justified.

This reconstruction of Hume's skeptical argument along the lines mapped out by Lipton has several interesting features. The argument is undeniably valid: if the premises are true, then the conclusion also must be true. So if all four premises are true, then the argument is sound. We shall focus on each premise in turn. Because of the wider epistemological issues it raises, we shall delay consideration of premise (1) until after we have looked at the other premises.

Premise (2) would be necessarily true if inductive arguments were simply defined as all those arguments that are not deductively valid. But this proposal would run counter to the way that we normally think and speak about inductive arguments. As indicated earlier, inductive arguments are usually regarded not simply as deductively invalid arguments, but as a proper subset of deductively invalid arguments. The class of inductive arguments thus includes universal and statistical generalization, statistical syllogism, induction to a particular, and so on, but it does not include the so-called deductive fallacies, such as denying the antecedent and affirming the consequent. Typically, deductively fallacious arguments are those in which the premises give no reason whatever for our believing that their conclusions are true. To keep the deductive fallacies at arm's length, then, we should replace the term *argument(s)* above in the first two premises with the more restrictive phrase "argument(s) in which the premises give (or are widely believed to give) a good reason for believing that their conclusions are true." Thus, premise (2) would now say arguments that are (or, are widely believed to be) reliable or truth-tropic are either deductively valid or inductive.

But even with this friendly amendment to premise (2), we are not quite out of the woods. The problem is that some forms of argument that are widely (but not universally) believed to be truth-tropic are neither deductively valid nor inductive. Arguments to the best explanation are a good example. This type of argument plays an important role in the debate over scientific realism, which is discussed in chapter 9. Some philosophers (such as Gilbert Harman) have gone so far as to claim that inductive

arguments are reliable only when arguments to the best explanation underlie and justify them.² But while it may be initially plausible to claim (with Harman) that "All observed As are B" is a good reason for believing that all As are B only because "All As are B" is the best explanation of why all observed As have been B, this sort of claim is not as plausible when it comes to other types of inductive argument, such as induction to a particular. For, surely, "The next A is a B" is not any sort of explanation of why all observed As have been B. One possible response to this objection would be to insist that IP is a reliable inference only because the conclusion that the next A is a B is derivable from the generalization that all As are B. The issue is a difficult one and cannot be resolved here. We shall simply say that the amended version of premise (2) would be true if all truth-tropic arguments that are not deductively valid are either explicitly inductive or ultimately reducible to inductive arguments. Alternatively, one could say that, insofar as premise (2) specifically concerns reliable arguments that can be used to justify IP (as opposed to inductive arguments in general), there is no nondeductively valid, noninductive argument (such as an argument to the best explanation) that can do the job.

Hume (and followers of Hume) assert premises (3) and (4) because they have a certain view of what it would take to justify an inductive argument such as IP. Obviously, IP is not deductively valid. The principle of inference that connects its premises with its conclusion is: from "All observed As have been B" and "The next individual is A" infer "The next individual will also be B." So to justify IP, we would have to show that this principle of inference is reliable. Hume argued that this principle and all the other principles of inductive arguments rest on a presupposition about the world, namely that patterns we have observed to hold in the past will continue to hold in the future; or without referring specifically to time, that as-yet-unexamined individuals will resemble those we have examined; or more generally still, that populations tend to resemble the samples that are drawn from them. Of course, these versions of what is often called the *principle of the uniformity of nature* are terribly vague. Lipton calls it the *principle of conservative induction* or the principle of "More of the Same" (416). But the vagueness does not matter to Hume's argument. For Hume reasons that no matter how, exactly, we spell it out, the principle of the uniformity of nature is an empirical assumption. Logic alone does not guarantee its truth. Hence, if we are to justify our belief in the uniformity principle by means of a deductively valid argument, then that argument must have among its premises another empirical statement that is at least as general as the uniformity principle we are trying to justify. So no valid deductive argument can yield the justification we seek, since it will contain a premise that is at least as controversial as the principle we are trying to justify. If we try to justify the uniformity principle inductively, then that argument in turn will presuppose the uniformity principle. Thus, our attempt at justification would fail on grounds of circularity.

Hume's argument for premises (3) and (4) is ingenious and controversial.³ Lipton gives a simple, intuitive gloss on premise (4)—the premise that denies that we can justify induction inductively—by comparing the principle of conservative induction, "More of the Same," with the principle of revolutionary induction, "It's Time for a Change." The revolutionary principle is counterinductive; it would license arguments of the following general form (CIP):

All observed As have been B.
 The next individual is an A.

 The next individual will *not* be a B.

Lipton invites us to imagine that someone were to offer as a reason for trusting the next application of the principle of conservative induction the argument that, since it has always (or nearly always) worked in the past, it should be trusted in the future. Lipton remarks that this inductive argument (of the form IP) cannot be a sound justification for conservative induction because the counterinductivist could offer an exactly similar argument (of the form CIP) for the conclusion that induction will *fail* in the future. The counterinductivist would argue that since conservative induction has always (or nearly always) worked in the past, by CIP, it will fail on its next application. Writes Lipton, "The justification of revolutionary induction seems no worse than the justification of conservative induction, which is to say that the justification of conservative induction looks very bad indeed" (417).

It is now time to consider the first premise of Hume's skeptical argument, which says that if IP can be shown to be justified, then there is an argument that shows it. This phrase about *showing* that IP is justified also appears in the conclusion of Hume's argument. Several epistemologists have recently emphasized the important differences among the following:⁴ a belief *being* epistemically justified, an argument *conferring justification* on its conclusion, and using an argument to show someone that a belief is justified. We have deliberately phrased Hume's argument in terms of the last of these three notions. To get a sense of how these notions differ, forget about induction for a moment and consider the analogous problem of trying to justify deduction.⁵ Let us focus specifically on the problem of justifying an elementary form of valid deductive argument, *modus ponens*:

If P then Q
 P

 Q

Most people can just see that *modus ponens* is valid without needing any argument to support that judgment. It is a self-evident truth, recognizable a priori, that *modus ponens* is necessarily truth preserving. For beliefs that we recognize as being justified by a priori intuition, no argument is needed, but we could, if we chose, supply an argument, the conclusion of which asserts that *modus ponens* is a valid form of inference. Inevitably, such an argument would itself be deductively valid and would thus rely on *modus ponens* or some other simple form of valid argument. Since this new argument is valid and, let us assume, its premises are true and justified, it confers justification on its conclusion. Though we did not need such an argument to convince us that *modus ponens* is valid, nonetheless the new argument transmits justification from its premises to its conclusion. The argument justifies the conclusion that *modus ponens* is valid.

Now suppose we are confronted with a deductive skeptic who wants us to prove to him that *modus ponens* is valid by giving an argument. It is pretty obvious that no argument could possibly meet his demand. Inductive arguments would be too weak; valid deductive arguments would, as far as he is concerned, beg the question, since they would presuppose the very thing—the validity of deductive inference—that he is not yet prepared to accept. Nonetheless, since the deductive argument we provide is valid and its premises are justified, it does justify its conclusion that *modus ponens* is valid. Thus, the fact that an inference is valid, and, moreover, that an argument can justify the claim that the inference is valid, is consistent with the impossibility of using an argument to show that this is the case to someone who is a skeptic about the general class of arguments to which the inference belongs.

The same moral holds for inductive inference. Even if Hume's skeptical argument is flawless, its conclusion is strictly limited. Hume has not proven that inductive inference is unjustified. Nor has he proven that no argument can justify the belief that inductive inference is justified. At best, he has proven that no one can use an argument to show an inductive skeptic that inductive inference is justified. When put in this way, Hume's conclusion is far less troubling than it might otherwise appear. For valid deductive inference is in the same boat with regard to the deductive skeptic, yet no one regards deductive inference as a philosophical scandal or even as being epistemologically suspect.⁶

THE PROBLEM OF DESCRIPTION

As Lipton points out, Hume's all-purpose skeptical argument allowed philosophers to raise the justification problem without having to consider in any detail how inductive inferences are actually made. The latter problem, the problem of description, has proven to be remarkably intractable. Lipton canvasses four solutions: the more-of-the-same model, the instantial

model, the hypothetico-deductive (H-D) model, and the causal model of Mill's methods.

The more-of-the-same model is not only extremely vague, but, Lipton argues, it also seems hard to reconcile with those inferences that predict change. He gives as an example the mechanic who predicts that the brakes on Lipton's car are about to fail. But the more-of-the-same model might still underlie the mechanic's inference about future brake failure if cars like Lipton's, with worn brake pads or leaking brake-fluid reservoirs, say, have invariably suffered brake failure in the past. Of greater difficulty for the more-of-the-same model are those inferences, common in the theoretical sciences, about hidden, perhaps unobservable, entities, drawn from observable data. Obviously these inferences cannot be understood merely as the extrapolation into the future of a pattern found to hold in the past.

Two responses suggest themselves. First, we could deny that inferences of this type are inductive and argue that they belong instead to another, noninductive category, such as arguments to the best explanation. Second, we might appeal to the distinction between the context of discovery and the context of justification, judging standard objections like those above as directed to the wrong context.

Giving a clear account of the distinction between the context of discovery and the context of justification is not easy, but the general idea is relatively straightforward.⁷ On the one hand, there is the psychological question about how a scientific hypothesis first arises in the mind of the individual scientist. This often involves inferences both deductive and inductive. But it also often includes nonrational influences as well as a liberal amount of inspired guesswork. Anything that helps answer the psychological question about the origins of scientific hypotheses falls within the context of discovery. On the other hand, once a scientific hypothesis has been formulated, questions about what kind of evidence supports it and to what extent place us in the context of justification, where we are concerned with relations of inductive support and confirmation. Many philosophers of science (especially Karl Popper) have argued that philosophers should confine their attention to the context of justification, leaving the context of discovery to psychologists, historians, and sociologists. Thus, a second response to Lipton's objection to more-of-the-same is to deny that a description of inductive inference should provide a literal, blow-by-blow account of how scientific ideas and hypotheses are generated in the minds of individual scientists. Rather, it should concentrate on describing the relations that hold between hypotheses and the evidence that confirms them in the context of justification.

Both the instantial model and the H-D model are intended by their proponents as descriptions of the inductive principles that scientists follow in the context of justification. In its most basic form, the instantial model simply says that generalizations are inductively confirmed by their positive instances. For example, the hypothesis "All As are B" is confirmed by

observed cases of *As* that are *B*. The usual objection to the instantial model is that it is both too weak and too strong. It is too weak because, as Goodman has argued, for any given piece of evidence there are just too many incompatible hypotheses that can claim that evidence as an instance.⁸ Also, the so-called raven paradox discussed by Hempel seems to show that on the instantial model utterly irrelevant evidence, such as a white shoe, must confirm the hypothesis that all ravens are black.⁹ The instantial model is judged too strong, since if it were a necessary condition for confirmation, far too many scientific theories, especially those involving unobservable entities and hidden causal mechanisms, could not be confirmed by observation.

The target of the last objection to the instantial model—its restriction to generalizations couched solely in the language of observation—naturally suggests the H-D model as a more realistic alternative. In its simplest formulation, the H-D model says that hypothesis *H* is confirmed by evidence *E* when *H*, in conjunction with some observation statements, entails *E* and *E* does not follow from those observation statements all by themselves. It places no limitation on the kinds of hypotheses that can be confirmed by evidence. But here again there are difficulties. Lipton mentions three. First, the H-D model is susceptible to the same grue and raven paradoxes as the instantial model. Second, the H-D model suffers from the problem of irrelevant conjunction (sometimes called the *tacking paradox*). For example, by conjoining “All crows are black” with “All swans are purple,” it seems that the H-D model would license the inference from the observation of a black crow to the prediction of purple swans.¹⁰ Third, by insisting that confirming evidence be entailed by any hypothesis that it supports, the model makes it impossible to confirm statistical hypotheses.

Many of our inductive judgments depend on inferring likely causes from known effects. The fourth account of induction discussed by Lipton focuses on the causal model encapsulated in Mill's methods. In his *System of Logic* (1843), John Stuart Mill (1806–73) proposed five “methods of experimental inquiry” for establishing causes. Lipton focuses on the first two of these methods, the method of agreement and the method of difference.¹¹

The Method of Agreement: If several instances of *P* have only one antecedent condition, *X*, in common, then *X* is the cause of *P*.

The Method of Difference: If two cases, one in which *P* occurs and the other in which *P* does not occur, have exactly the same antecedent conditions with one exception, namely, that *X* is present in the first but absent in the second, then *X* is the cause (or a necessary part of the cause) of *P*.

Both these and the other methods that Mill proposes are best regarded as applying to inferences in the context of justification rather than the context of discovery. In the discovery phase, we must first isolate the genuine cases of *P* and then identify all the factors (*A*, *B*, *C*, etc.) in each of these cases that might be causally relevant to *P*. Only after this hard work has been done can we use Mill's methods to eliminate rival hypotheses about the cause of *P*. This emphasis on elimination is a distinctive feature of Mill's approach to induction. The previous three models—more-of-the-same, the instantial model, and the H-D model—all regard inductive inference and the confirmation of hypotheses as proceeding by accumulating evidence for a particular hypothesis; according to Mill's methods, induction and confirmation work by eliminating rival hypotheses.

The method of agreement is designed to identify a necessary condition for *P*. In other words, it is intended to isolate a factor without which *P* could not occur. But in reality things can be more complicated than Mill's method suggests. For example, it may be true that, in four cases of *P*, the only single factor they all had in common was *A*. But that does not guarantee that *A* is necessary for *P*. It might be that two of the cases involved *B*, and the other two cases involved *D* so that the complex condition (*B* or *D*) is what is necessary for *P* and that *A* is in fact irrelevant. Moreover, the method of agreement says nothing about sufficient conditions for *P*. Again, *B* and *D* might each be sufficient to produce *P*, and *A* could be irrelevant.

The method of difference complements the method of agreement by describing a procedure aimed at identifying a sufficient condition for *P*. But as Mill himself realized, even when only one difference exists between a pair of cases, it does not follow that this factor is sufficient to produce *P* all by itself. For the distinguishing factor might, as Mill recognizes, merely be a necessary part of a sufficient condition for *P*. For example, suppose that the only difference between Adam (who gets ill) and Bob (who stays well) is that Adam was exposed to the disease organism, *D*. It does not follow that exposure to *D* will always produce illness in anyone exposed to it, for the illness might also depend on the person in question lacking immunity to *D*. All that we are entitled to assert is that *D* is probably a necessary part of a complex of factors that are jointly sufficient to produce *P*.

The limitations of Mill's two methods as a descriptive model of inductive inference are fairly obvious. At best they apply to cases in which all the factors of potential causal relevance are observable and can be enumerated prior to the application of Mill's methods. The vast majority of scientific cases are not like this, either because the theory to be confirmed is not causal or because the theory invokes entities and processes that are not observable. Furthermore, as Lipton points out, the antecedents of Mill's two methods require a single agreement or a single difference

respectively. This condition is totally unrealizable in actual cases, where thousands of agreements and differences could be identified. Much of the real work in such cases is done by picking out from this vast array the relatively small number of factors that might be causally relevant. Since this selection is made before Mill's methods come into play, it means that Mill's methods do not capture important parts of the pattern of inductive inference.¹²

4.2 | Popper's Rejection of Induction

Sir Karl Popper (1902–94) is the most influential anti-inductivist philosopher of science of the twentieth century. Educated in Vienna during the heyday of logical positivism, Popper was one of the earliest and severest critics of the logical positivists' verifiability criterion of meaning and their attempt to use verifiability (and, later, confirmability) as a demarcation criterion between science and pseudoscience. Rejecting the positivists' criterion, Popper developed and proposed falsifiability (not verifiability or confirmability) as his own demarcation criterion. (See chapter 1 for a discussion of Popper's falsificationist demarcation criterion and its shortcomings.)

In "The Problem of Induction," Popper argues that there is no such thing as inductive confirmation. No scientific theory or law is made probable by evidence. Rather, science follows the method of conjectures and refutations. Scientists propose bold conjectures and then devise severe tests to falsify them if, indeed, they are false. A theory that withstands such tests is not confirmed or made probable by its successful predictions. Instead, it is what Popper calls *corroborated*. This, in very brief outline, is Popper's anti-inductivist falsificationist philosophy of science.

WHY INDUCTION CANNOT BE JUSTIFIED

Popper's case against induction is essentially Hume's, and Popper thinks it is unanswerable. There are just three small differences between Popper's skeptical argument and our reconstruction of Hume's argument in the preceding section of this commentary. First, instead of saying that any attempt to justify induction inductively must necessarily be circular, Popper says that it must lead to an infinite regress of inductive principles of ever higher order. Popper's idea is that when trying to justify a particular inductive principle or inference, we might avoid begging the question by employing an inductive principle or inference that is different from the one we are trying to justify. But this principle or inference, in turn, will be synthetic and hence will need to be justified by experience, thus requiring a further inductive argument, and so on. Popper assumes that such

an infinite regress is vicious because the first inference will not be justified unless the second one is, and the second one will not be justified unless the third one is, and so on. Since the chain of justifying arguments can never be completed, the first inference never receives the justification it needs.

A second difference between Popper's argument and Hume's is that Popper considers, but nevertheless rejects, Kant's attempt to accord the principle of induction (in the guise of the "principle of universal causation") the status of an a priori synthetic truth. (Kant made his proposal after Hume's death.) If knowable a priori as Kant claimed, the causal principle would not need to be justified on the basis of experience. Indeed, Kant credited Hume, especially Hume's skeptical argument about induction, with having wakened him from his "dogmatic slumbers," from his uncritical assumption that the fundamental laws and principles of science could be justified solely on the basis of experience and perception.

Given the obscurity of Kant's arguments in the second Analogy of his *Critique of Pure Reason*, Popper wisely avoids giving them any detailed consideration in our selected reading. But elsewhere, Popper has criticized Kant's position in the following way.¹³ Popper agrees with Hume and Kant that our belief in the uniformity of nature cannot be justified from the bottom up by making ever wider generalizations from the regularities we discover in nature because every one of these inferences already presupposes the point at issue, namely the uniformity of nature. Popper also agrees with Kant that the lawlikeness of nature must be imposed by us from the top down (and thus is prior to experience). But Popper sees no reason for Kant's rationalist confidence that our search for particular regularities must succeed or that our belief in those regularities can be justified a priori. According to Popper, Kant was misled by the tremendous success of Newton's mechanics into believing that certain high-level generalizations about nature (such as those encapsulated in Newton's laws) must be true and that our search for such regularities must succeed. But as Popper puts it, all our knowledge is conjectural. Kant had confused the alleged psychological indispensability of certain of our beliefs about nature with an a priori guarantee that specific versions of those beliefs must be true.

A third wrinkle in Popper's argument that distinguishes it from our reconstruction of Hume's reasoning actually involves a point that Hume himself made. Hume anticipated that some inductivists (as we shall call them) might concede that we cannot prove with certainty that the conclusion of any inductive argument must be true. But, they might argue, we can nonetheless show that these conclusions are likely or highly probable relative to the premises from which they are inferred. Popper follows Hume in dismissing this strategy as unavailing, charging that this probabilistic version of inductivism leads either to "apriorism" or to a vicious infinite regress.

WHY THERE IS NO LOGIC OF DISCOVERY

Popper's main concern in section 2 of our reading, "Elimination of Psychologism," is to distinguish descriptive questions from normative ones. Empirical psychology, he says, may investigate how and why a new theory suggested itself to a scientist, but tracing this causal history is irrelevant to evaluating the theory. Only when the process of conceiving a new idea has been completed can the normative appraisal of what has been conceived begin. After the new theory has been formulated, its subsequent evaluation may in some cases mirror the actual thought processes of scientists and in that sense provide a rational reconstruction of them, but the epistemological evaluation is not intended as a psychological description and should not be judged by how well or ill it happens to reflect actual patterns of reasoning. Moreover, he concludes that there is no set of rules, no logical method for generating ideas. He agrees with Bergson and Einstein that scientific discoveries always involve an element of creative intuition that cannot be captured in any algorithm.

Despite the frequency with which this short section from *The Logic of Scientific Discovery* is cited, it is difficult to get straight what Popper is claiming and what his arguments are. Popper seems to be making two main points: that empirical psychology and history are irrelevant to epistemology and that there is no interesting version of a logic of discovery. These two points are independent, and neither depends on Popper's anti-inductivism. So, for example, inductivists such as Reichenbach and Hempel can endorse both positions while still rejecting Popper's inductive skepticism. Also, someone can accept Popper's first point (by insisting that epistemology cannot be naturalized) and yet insist that heuristic rules and maxims can guide the construction of new theories. All that anti-inductivist falsificationists (such as Popper) and inductivist hypothetico-deductivists (such as Hempel) rule out is the possibility of an algorithm for generating new theories that would guarantee the truth or high probability of those theories. But the rejection of such an infallibilist, mechanical conception of a logic of discovery leaves ample room for more modest versions, and not even Popper in his most outspoken moments would deny that reasoning (*deductive* reasoning) can play a role (however fallible and justificationally impotent) in the generation of new theories.¹⁴

FALSIFICATIONISM AND CORROBORATION

According to Popper the logic of science is purely deductive. The essence of rational scientific method is not to look for probable theories and try to confirm them with evidence, but to invent bold, improbable theories and try to refute them. Theories are tested by deducing predictions from them. If the predictions are false, then it follows deductively (by *modus tollens*) that the theory is false and should be discarded. If the predictions

are verified, then the theory is not confirmed but corroborated. ("Being corroborated" and "having a high degree of corroboration" are terms of positive epistemic appraisal for Popper.) Withstanding a severe test proves a theory's mettle. Corroborated theories should not be accepted as true or even as probable, but nonetheless it is rational for us to rely on them at least until better theories come along. Eventually, by continually replacing falsified theories with ever bolder ones that have not yet been refuted, science progresses towards the truth. Or so Popper claims.

There are many problems with Popper's falsificationism as an account of scientific rationality. One problem is reconciling Popper's talk of "verifying" the predictions deduced from theories with his insistence that no empirical statement whatever can be established as true or highly probable. Popper's response is to say that predictions, in the form of "basic statements," are accepted by the conventional decision of the scientific community and that these decisions can be reversed later.¹⁵ Popper adamantly denies that any basic statement is made probable by being grounded in or supported by the available evidence. Scientists simply make a free, unconstrained decision to accept the statement, at least until further notice. In a memorable and oft-quoted passage, Popper writes:

The empirical basis of objective science has thus nothing "absolute" about it. Science does not rest upon rock-bottom. The bold structure of its theories arises, as it were, above a swamp. It is like a building erected on piles. The piles are driven down from above into the swamp, but not down to any natural or "given" base; and when we cease our attempts to drive our piles into a deeper layer, it is not because we have reached firm ground. We simply stop when we are satisfied that they are firm enough to carry the structure, at least for the time being.¹⁶

But as many of Popper's critics have argued, this metaphor of the swamp and the doctrine it illustrates are deeply troubling, for despite Popper's claim that conventional acceptance occurs only at the level of basic statements, not at the level of theories, surely the refutation of a theory must be just as conventional as the basic statement that prompts it. Thus, it is hard to see how Popper can justifiably assert that science is rational and objectively progressive when it ultimately depends on purely conventional and arbitrary decisions.¹⁷

4.3 | Salmon's Criticism of Popper's Anti-Inductivism

As the preceding discussion has revealed, Popper's anti-inductivism creates difficulties for falsificationism by making the ungrounded acceptance of

basic statements (and thus the refutation of theories) seem arbitrary and irrational. In his article, "Rational Prediction," Wesley Salmon also takes aim at Popper's rejection of induction. Salmon charges that, without induction, Popper cannot give a coherent justification for using corroborated theories to make what Salmon calls *rational predictions*.

THE PROBLEM OF RATIONAL PREDICTION

Although Popper uses the term *prediction* to refer exclusively to the consequences drawn from a theory in order to test it, Salmon reminds us that there are at least two other common reasons for making predictions. Sometimes, predictions are made solely to satisfy our intellectual curiosity about what will happen in the future. These predictions are of purely theoretical interest. More frequently, predictions are made not merely to satisfy our curiosity but because something of practical importance is at stake. For example, we might need to know how a high-rise apartment building will behave during a moderate earthquake to make sure that its design is safe. We could engage in blind guesswork or turn to clairvoyants for help, but the best guide is to rely on currently accepted scientific theories.

Salmon claims that, because Popper rejects any form of inductive confirmation, he is unable to do justice to rational predictions. (For the purposes of his article Salmon focuses on those rational predictions that are ingredients in the making of practical decisions.) If we are never justified in accepting any theory, and successful predictions and explanations can never count as confirming evidence for a theory, how can falsificationists such as Popper and his defender, John Watkins, explain why current scientific theories are a better guide to the future than blind guesswork, clairvoyance, or reading Chinese fortune cookies?

Popper and Watkins have responded to Salmon's challenge by appealing to the notion of corroboration. A theory is corroborated at a given time when it has survived our attempts to refute it up to that time; one theory has a higher degree of corroboration than another when it has survived more severe tests. Suppose we have to choose a theory as the basis for a practical prediction about the future from among a number of rivals none of which has yet been refuted. In this situation, it seems only rational to prefer the theory with the highest degree of corroboration. For in this kind of situation, Watkins says, we have nothing better to go on. Popper and Watkins argue that selecting in this way is not tantamount to making an inductive inference because judgments about corroboration are based solely on a theory's past performance: they say nothing and imply nothing about how well the theory will do in the future. Moreover, in choosing the best corroborated theory we are not "relying" on that theory in the sense of accepting it as true or even probable. We are simply making a choice.

THE INADEQUACY OF POPPERIAN CORROBORATION

Salmon finds the Popper-Watkins “solution” to the problem of rational prediction completely unsatisfactory. In a tone of patient but growing exasperation with the falsificationists, Salmon makes two main points. First, when trying to justify our selection of a theory for the purposes of practical prediction, it is not literally true to say, as Popper and Watkins do, that we have nothing else to go on. There are plenty of other ways to choose a theory. We could, for example, choose the theory which is shortest when written out in English or the one that gets the most votes on a TV call-in show. Presumably, what the falsificationists mean is not that there is nothing else to go on but that there is nothing else *rational* to go on. Yet even if all other methods are irrational, this does not entail that choosing the best corroborated scientific theory is rational. For it is perfectly possible that all methods are without exception equally irrational. Thus, in saying that there is nothing else to go on, Popper and Watkins must actually mean that there is nothing else to go on that is rational apart from corroboration. And that, Salmon complains, simply begs the question.

Second, Salmon sees an irreconcilable conflict between Popper's insistence that corroboration statements have no predictive import and his assertion that they justify our preference for one theory over another for purposes of rational prediction. If corroboration statements are solely about the past, how, without tacitly making an inductive inference, can they justify the choice of a theory on which to base predictions about the future?¹⁸

4.4 | Hempel on Criteria of Confirmation

In “Criteria of Confirmation and Acceptability,” excerpted from his lucid introduction to the philosophy of science, *Philosophy of Natural Science*, Carl Hempel surveys the factors that scientists use in deciding whether a hypothesis is acceptable or probably true. These factors include inductive confirmation by the available evidence, simplicity, and support from other theories. Hempel attempts to clarify these factors, but unlike, for example, the Bayesian approach to scientific inference (discussed in chapter 5 of this book), he does not offer a systematic, unified theory to justify them. At the end of his article he briefly discusses the difficulties of providing such a theory within the framework of Carnap's system of inductive logic. Hempel presupposes that, because they involve universal generalizations, scientific hypotheses and laws cannot be conclusively verified, no matter how extensive the evidence in their favor. But like most other philosophers of science, Hempel rejects Popper's falsificationism and accepts that sci-

entific hypotheses and laws can be made more or less probable by evidence.

QUANTITY, VARIETY, AND PRECISION OF EVIDENCE

As a general rule, the more evidence there is for a hypothesis, the more strongly the hypothesis is confirmed. But as Hempel notes, there is a law of diminishing returns with regard to evidence. Repetitions of the same kind of experiment or observation, under the same conditions, rapidly lose their power to add substantially to the credibility of the hypothesis being tested, even when all the observations are exactly as the hypothesis predicts. What we desire in evidence is not mere quantity, but diversity. Different kinds of test, under widely varying conditions, have a greater power to confirm than do repetitions of the same kind of test under the same conditions. More generally, variety in evidence makes for strength in confirmation. Why is this?

Hempel argues that when the evidence is diverse, more of the hypothesis gets tested than when the evidence is narrow. He gives as an example, Snell's law in optics. For convenience, Hempel lets S stand for Snell's law, the hypothesis that, for any pair of transparent media, the ratio of the sine of the angle of incidence (i) to the sine of the angle of refraction (r) is a constant. This is a generalization about all associated angles of incidence and refraction, and all pairs of transparent media. Thus, confining our observations to air and water and to angles of incidence of 30 degrees would leave a large part of S untested. Let us call this restricted test, T_1 . By varying our experiments to include a range of angles (test T_2) and many different media (T_3), more of S would be tested. But why should evidence collected by testing more of S lead to greater confirmation, assuming that the total number of data points—the quantity of the evidence for S —remains unchanged?

Hempel's answer appeals to falsifiability. Let S_1 be the hypothesis that the ratio of $\sin i$ to $\sin r$ is a constant for all samples of air and water when angle i is 30 degrees. S implies S_1 , but S_1 could be true and S false, and the restricted test T_1 would not show it. In other words, if S_1 were true, T_1 could not falsify S . Similarly, if S_2 is the hypothesis that the ratio of $\sin i$ to $\sin r$ is a constant for all samples of air and water and all associated pairs of angles, then if S_2 were true, T_2 could not falsify S . Thus, Hempel concludes, the more thorough the test (that is, the greater the power of the test to falsify a hypothesis), the more support a favorable outcome gives to the hypothesis. More diverse evidence results from more powerful tests; the more powerful the test, the greater the confirming power of the evidence it generates.

Hempel's explanation of the power of diversified evidence gives an inductive twist to Popper's notion of the falsifying power of a test. But

Hempel never spells out in detail why greater falsifying power implies greater confirming power. One way of connecting the two would involve the Bayesian approach to confirmation explored in chapter 5. For, as Hempel notes, we value diversity not for its own sake, but because we suspect it might affect the results of our experiments. To the Bayesian, diversity has evidential significance only when it is associated with competing hypotheses that have some initial degree of likelihood or probability. By disconfirming or falsifying some of these competitors, powerful tests thereby increase the probability of the surviving hypothesis (or hypotheses).

Hempel's discussion of the value of precision in evidence is brief. Increased precision of measurement has special significance when we are testing hypotheses that make identity claims (such as the identity of inertial and gravitational mass postulated by Einstein's general theory of relativity) or which predict a null result for a class of measurements (as does the special theory of relativity for all attempts to measure absolute velocities or to ascertain which inertial frames are at rest with respect to absolute space). For in these cases, the level of precision at which the results in question are supposed to hold is unlimited.

NOVEL PREDICTION

Hempel does not use the phrase *novel prediction*. Instead, he writes of "new" evidence and "new" phenomena. The word *new* is in quotes because the evidence or phenomena need not be new in the sense of not having been previously recorded or observed. Consider, for example, the case of the Swiss schoolteacher Johann Jakob Balmer (1825–98), discussed by Hempel. On the basis of four observations made by the Swedish physicist Anders Jonas Ångström (1814–74), Balmer published, in 1885, a general formula giving the wavelengths of a series of lines—now called the Balmer series—in the visible range of the emission spectrum of hydrogen. The original data set on which Balmer based his formula was already widely known by scientists at the time. Balmer then used his formula to predict further lines in the hydrogen spectrum, which were subsequently verified, thus confirming his formula. As Hempel notes, some of the additional lines had in fact already been observed in 1885, when Balmer put forward his theory, but Balmer and most other scientists were unaware of this fact. Thus, to call the later predictions new or novel does not mean that no one had previously observed them but rather that the results of these observations, were not widely known. In particular, the results were not known to Balmer at the time he proposed his formula.

In analyzing the debate over novel prediction and confirmation, it is helpful to distinguish among the various senses in which a prediction might be considered novel. We shall introduce the phrase *epistemic novelty* to distinguish the restricted epistemic sense of predictive novelty—not widely known and not known by the person proposing the theory—from

the wider sense of *temporal novelty*—not known by anyone at the time the theory was proposed. Obviously, both senses of *novelty* are epistemic and relative to time, but epistemic novelty is also relative to persons: even if two scientists were to propose the same theory at the same time, the theory could make predictions that were epistemically novel for one of the scientists, but not for the other, depending on what they knew and when they knew it.

Some philosophers of science have also deemed two further senses of *novelty* important. We shall call these *design-novelty* and *use-novelty*. A result is design-novel so long as the scientist did not deliberately construct the theory to yield the result in question. For example, when Newton devised his gravitational theory, he was perfectly well aware of the precession of the equinoxes, but this was not an explicit constraint on his theorizing. He did not, for example, go through several preliminary versions of his theory, rejecting those that did not give the right answer for the rate of precession. Contrast this with Einstein, who did reject earlier versions of the general theory of relativity because they failed to imply the correct value for the rate of precession of the perihelion of Mercury. Design-novelty is what Elie Zahar and others have called *heuristic novelty*: a result is novel in this sense if it “did not belong to the problem-situation which governed the construction of the hypothesis.”¹⁹ Use-novelty is a special case of design-novelty. A prediction lacks use-novelty when that prediction is used to fix the value of a free parameter in a theory or when the result in question is simply built into the theory. (The relevance of this sense of novelty for confirmation has been emphasized by John Worrall.²⁰)

There are simple logical relations among the four notions of novelty that we have distinguished: temporal novelty implies epistemic novelty, epistemic novelty implies design-novelty, and design-novelty implies use-novelty; but use-novelty does not imply design-novelty, design-novelty does not imply epistemic novelty, and epistemic novelty does not imply temporal novelty.

Several different questions can be raised about novel predictions and confirmation. Here are two of the most important:

If a theory is deliberately designed to accommodate or explain phenomena that are already known, do those phenomena confirm the theory that predicts them? (If *T* is designed to explain *E*, does *E* confirm *T*?)

Other things (such as quantity, diversity, and precision of evidence) being equal, does a theory derive greater confirmation from novel predictions than from the already-known phenomena that it explains or accommodates? (Is *T* confirmed more strongly by novel prediction *D* than it is by *E*?)

Since there are two questions and at least two possible answers to each (Yes, No, It all depends), at least four positions are possible. In practice, five positions have been held by significant numbers of scientists and philosophers of science. The most extreme position would answer "No," emphatically, to both questions, thus denying any form of inductive confirmation by evidence. This is Popperianism, which was discussed in the preceding sections of this commentary. Few philosophers of science regard this kind of complete rejection of inductive confirmation as plausible.

Two other extreme positions (but less extreme than Popperianism) are often referred to as *explanationism* and *predictionism*. Explanationism is the view that only the explanation of previously known and accepted results can confirm a theory and that novel predictions have no power to confirm. Thus the explanationist answers the first question with an emphatic "Yes" and the second question with an emphatic "No." Something close to this view has been defended by Stephen Brush and John Worrall. Predictionism is the opposite view that only novel predictions can confirm a theory and that the explanation of previously known and accepted results has no power to confirm. Thus predictionists respond to the first question and the second question with an emphatic "No" and an emphatic "Yes," respectively.

Between the two extremes of explanationism and predictionism are two more conciliatory positions. One popular choice is to answer "Yes" to both questions. This position, which we shall call *accommodationism*, allows that the explanation or accommodation of old evidence has some confirmatory value but insists that novel prediction must always have more. The other position, which we shall call *formalism*, agrees that accommodating old evidence can confirm but holds that, other things being equal, novel predictions are no better and no worse than old ones. Whether or not a particular novel prediction confirms better than an old one depends on its evidential content, not on its novelty. Formalism is driven by the insistence that confirmation depends solely on the formal, inductive, logical relation between propositions, regardless of when and by whom those propositions were known or to what purposes they were put. Our taxonomy of positions on the novel prediction debate is summarized in table 1.

Notice how explanationism, accommodationism, and formalism differ from each other. Explanationists deny that novel predictions are superior to explanations because they deny that novel predictions have any power to confirm. Accommodationists accept that novel predictions can confirm and think that the confirmation they provide is always superior to that of explanations. Formalists think that novelty per se is irrelevant: all that matters is the evidential content of the propositions involved, whether they be predictions or explanations. Sometimes novel predictions confirm better than explanations, sometimes they do not; it all depends on the propositions involved.

	Question 1: If T is designed to explain E , does E confirm T ?	Question 2: Is T confirmed more strongly by novel prediction D than it is by E ?
Popperianism	No	No
Explanationism	Yes	No
Predictionism	No	Yes
Accommodationism	Yes	Yes
Formalism	Depends	Depends

Table 1

Hempel's answer to the first question is "Yes." Without elaborating, he simply asserts that it follows from the fact that theory T entails old evidence E that E must confirm T . Without placing further qualifications on the relationship between T and E , this assertion seems dubious. For example, if we were to construct a theory by conjoining quantum mechanics, Q , with the proposition E (which we already know to be true), most people would judge that the resulting theory ($Q \& E$) is *not* confirmed by its true entailment, E . As the critics of the H-D model are fond of pointing out, entailment of true consequences is too weak to be a sufficient condition for confirmation.

With regard to the second question Hempel notes that most people think that the confirming power of novel predictions is greater than that of explanations or the accommodation of old evidence. He then raises "a puzzling question" (450). What if Balmer's equation, T , had been constructed only after the new evidence, D , was in, so that T was designed to fit all the evidence ($E \& D$)? That T fits the evidence under these circumstances is hardly surprising, which seems to support the predictionist preference for new evidence. But, Hempel replies, something in this case is still surprising. The surprise is not that T fits ($E \& D$), for with sufficient ingenuity, numerous theories can be invented that would fit any given evidence. But most of those theories would be horrendously complex and arbitrary. No, the surprise is that the theory, T , which fits ($E \& D$) is such a simple mathematical formula. Thus, we would still have a reason for thinking that T is true, even though T was tailored to fit the evidence. What confirms T is not the fact that T fits ($E \& D$) but that T is quite simple. Still, Hempel does not endorse this simplicity argument as a rejoinder to the predictionist because its plausibility depends on whether we

can connect simplicity with truth, and that, in turn, depends on whether we can give an adequate objective characterization of simplicity—an issue Hempel takes up later in his article.

Hempel concludes this part of his discussion by noting that, quite apart from the arguments he has been exploring, from a logical point of view all that matters is the content of theory T and the evidence E , not when they were proposed or discovered. Again, he is not taking sides on the novel prediction issue but merely noting that if one adopts the inductive-logical approach to confirmation of Carnap and others, then the formalist view naturally follows. Deductive logic provides an appropriate analogy. Just as the deductive validity of an argument depends solely on the content of its premises and conclusion, so, too, does the inductive strength of an argument. The inductive probability of the conclusion given the premises is an objective, logical matter that is entirely unaffected by how the theory was constructed or when the truth of the premises was discovered.

THEORETICAL SUPPORT

Not all support for a theory or hypothesis comes from “below,” from inductive confirmation by successful predictions; some comes from “above,” from other well-established and well-accepted theories. The plausibility of new theories and hypotheses is often assessed by how well they stand up in the light of evidentially well-confirmed theories that we already accept. The strongest relation between one theory and another is deductive. Hempel gives as typical examples of deductive support from above the support that Balmer’s formula received when a more general formula was confirmed of which Balmer’s expression was a special case and the support that these equations received in turn from Bohr’s well-confirmed theory of the hydrogen atom that explained them. Hempel uses the phrase *deductive support* in discussing these cases because of the deductive relation between the theories involved. But the support that one theory confers on another is still inductive. The basic idea is that if a theory, G , is logically derivable from some more general theory H and H is well-confirmed by a range of evidence, E , other than the evidence that confirms G , then G is also confirmed by E . Elsewhere in his writings on confirmation theory, Hempel has called this principle the *special consequence condition*:²¹

If E confirms H , and H entails G , then E confirms G .

The special consequence condition says, in effect, that whatever confirms a hypothesis H also confirms any statement that is deducible from H .²² Imagining any confirmation theory that could dispense with this principle is difficult, since we presuppose it whenever we infer that the predictions of a well-confirmed theory are likely to be true. A similar looking

principle that Hempel named the *converse consequence condition* also has an air of plausibility:

If *E* confirms *G*, and *H* entails *G*, then *E* confirms *H*.

Newton's gravitational theory is commonly regarded as being confirmed in virtue of its entailing (with certain additional hypotheses) Galileo's law of falling bodies and Kepler's laws of planetary motion. Galileo's law and Kepler's laws were well confirmed before Newton advanced his theory. So when Newton was able to derive these laws from his new theory, the theory immediately inherited a substantial measure of inductive confirmation. Indeed, on at least one model of scientific progress according to which scientific knowledge grows by a process of theoretical reduction, something like the converse consequence condition would seem to be an integral part of the logic of confirmation. (For an extended account and criticism of this model, see chapter 8.)

But as Hempel has shown, the special consequence condition and the converse consequence condition together have the unacceptable, in fact absurd, consequence that any observation report confirms any hypothesis whatsoever. To show that we cannot accept both conditions as principles of confirmation, consider the following example. Let *E* be the observation report, "This is a black crow." *E* confirms the hypothesis, *G*, that all crows are black. Let hypothesis *H* be the conjunction of two generalizations, *G* and *J*: all crows are black and all swans are purple. Trivially, *H* entails *G*. Therefore, by the converse consequence condition, *E* confirms *H*. But, equally trivially, *H* entails *J*. So by the special consequence condition, *E* confirms *J*. Thus, by employing both conditions, we reach the absurd consequence that "This is a black crow" confirms "All swans are purple."

The special consequence condition appears to be irreproachable. For, whatever the connection between probability and confirmation, any proposition must be at least as probable as the propositions from which it can be derived. Thus, it seems inevitable that inductive confirmation should be transmitted downwards to the logical consequences of any confirmed proposition.²³ Since we cannot embrace both conditions, we must give up the converse consequence condition. And this is the course that Hempel takes in his formal study of confirmation theory. Hempel thinks that the plausibility that the converse consequence condition seems to derive from classic cases of theory reduction is spurious. For in the classic cases, not merely are the well-established laws logically derivable from a more general theory, but, crucially in Hempel's view, the laws are *instances* of the more general theory. In his formal theory of confirmation, Hempel advocates the *satisfaction criterion*: a generalization, law, or theory is inductively confirmed by an observation report only if the report is an instance of the statement it confirms.²⁴ We cannot pursue the details of Hempel's

theory here. Suffice it to say that it represents a significant departure from the crude form of the H-D model. For according to Hempel's satisfaction criterion, hypotheses, theories, and laws are no longer confirmed merely by the true predictions that can be deduced from them. For a theory to be confirmed, those predictions must be not merely logical consequences but instances of the hypotheses, theories, and laws in question. Thus, while "all crows are black" is confirmed by the report of a black crow, Hempel's theory denies that "this is a black crow" confirms "all crows are black and all swans are purple" because the report "this is a black crow" is not an instance of the conjunctive hypothesis. In simple terms, Hempel accepts that whatever confirms a theory also confirms any logically equivalent or weaker theory, but he denies that whatever confirms a theory must also confirm every stronger theory.

SIMPLICITY

The notion of simplicity has already been invoked in the section of Hempel's article on novel predictions. There, in response to the predictionist claim that a theory cannot be confirmed by evidence it was designed to explain, Hempel suggested a reply that assumed that simpler theories are more likely to be true than more complex ones, regardless of how they have been arrived at. Hempel now examines whether this and similar claims about simplicity can be justified.

Like most philosophers who discuss simplicity in scientific theories, Hempel takes curve fitting to be a paradigm illustration. In Hempel's example, the four data points are the four pairs of values of the variables u and v —(0, 2), (1, 3), (2, 4), and (3, 5)—to which we wish to fit a curve. Many different hypotheses are consistent with these data. Intuitively, the simplest of them is the straight-line hypothesis $v = u + 2$ that Hempel calls H_1 . But many other hypotheses expressing v as a function of u are possible—hypotheses involving additional terms and higher powers of u . In cases like this, the philosopher's task is twofold: first, to explicate the concept of simplicity so that hypotheses such as H_1 are judged correctly to be simpler than their rivals; second, to justify what Hempel calls the *principle of simplicity*: that, given two otherwise equally well-confirmed rival hypotheses, we should prefer the simpler; in other words, that simplicity enhances the credibility of a theory or hypothesis.

Hempel briefly surveys some of the attempts to analyze the notion of simplicity and the difficulties such attempts have encountered. A guiding assumption in this search has been that, whatever it ultimately amounts to, simplicity should be objective, not simply a matter of familiarity, convenience, or intuitive appeal that can vary from person to person or that is relative to context. Thus, merely trying to judge how complicated algebraic equations appear to be is a poor way of measuring simplicity. So, too, is looking at the graphical representation of a curve in some coordi-

nate system. In both cases, the judgment made will be relative to the variables or the coordinate system chosen. In Cartesian coordinates (x, y) , a straight line will appear simple and a spiral complex; in polar coordinates (r, θ) the spiral will appear simple and the straight line complex. An alternative might have us counting axioms and judging that theory to be simplest whose axioms are fewest. But like the previous proposals, this fails due to arbitrariness: one can always reduce the number of axioms by conjoining statements.

Hempel discusses and criticizes four attempts to justify the simplicity principle. The first two are based on ideas that have been held by a number of scientists and philosophers; the last two derive specifically from Reichenbach and Popper, respectively.

■ THE SIMPLICITY-POSTULATE ARGUMENT

The simplicity postulate asserts that most of the fundamental laws of nature are simple. If we knew that to be true, then we would have an obvious rationale for embracing the principle of simplicity, which holds that simpler theories should be preferred because they are most likely to be true. The problem, as Hempel observes, is that we do not know that the simplicity postulate is true. In fact, the assumption that the laws of nature are simple is at least as controversial as the principle of simplicity itself and, hence, cannot be used to justify it without a great deal of further argument. Merely appealing to our past success in uncovering some relatively simple laws of nature will not suffice here, since the relevant question is precisely, Why should this past success be taken as good evidence that the laws of nature as a whole are simple? Obviously, it would be circular to appeal to the simplicity of the simplicity postulate as a reason for preferring it to more complex postulates.

One of these alternative postulates is that a principle of selection is driving us to find the simple laws and that the laws we have discovered thus far are not a representative sample of the laws of nature as a whole. Rather, the laws we have discovered are simple because those are the ones on which we have chosen to focus our attention. Both the simplicity postulate and the selection hypothesis explain the evidence, namely, the relative simplicity of the laws uncovered by science thus far. So what reason can we give for preferring the former to the latter that avoids begging the question?

■ THE ECONOMY-OF-SCIENCE ARGUMENT

In the generation of scientists and philosophers of science preceding the advent of logical positivism, positivistically inclined thinkers such as Ernst Mach, Karl Pearson, and Pierre Duhem insisted that the principal aim of science is not to explain but to find the most economical description of facts. On this view, theories are not hypotheses about the structure of reality; they are merely tools for summarizing information obtained from

observation and experiment. Hence, the simplest theories should be preferred because they provide the most succinct, shorthand summary of the available data.

The flaw in this attempt to justify the simplicity postulate was exposed by Hans Reichenbach (among others) in his book, *Experience and Prediction*, and Hempel's remarks in his article echo Reichenbach's criticism. Reichenbach introduced the phrases *descriptive simplicity* and *inductive simplicity* to stand for two very different notions of simplicity. Descriptive simplicity concerns empirically equivalent theories where the choice of one theory over another is, to quote Reichenbach, "nothing but a matter of taste or of economy."²⁵ Reichenbach gives several examples: the preference of scientists for the metric system over older systems for making measurements and performing calculations; the choice of one particular inertial system as the "rest system" in the special theory of relativity and the adoption of a particular definition of simultaneity in that theory; the decision to adopt Euclidean rather than non-Euclidean geometry as the geometry of space. But when, as in Hempel's illustration, we are trying to fit curves to a finite set of data points, the rival hypotheses are not empirically equivalent. They are hypotheses whose predictions about future measurements disagree. In curve-fitting problems, the issue is not, which hypothesis gives the most economical description of the facts already observed? but rather, which hypothesis is the most reliable guide to observations that will be made in the future? In short, inductive simplicity concerns what Wesley Salmon calls *rational prediction*, not economical description. Reichenbach (and Hempel) accuse positivists like Mach of confusing the two concepts of simplicity. Economy and convenience govern the choice between (equivalent) theories that differ merely in descriptive simplicity. But when (nonequivalent) theories differ in inductive simplicity, the choice depends on which theory makes the best predictions about the future.

■ REICHENBACH'S INDUCTIVE ARGUMENT

Reichenbach's defense of the principle of simplicity is an application of his well-known pragmatic vindication solution to the problem of induction. Whether one chooses the smoothest curve that fits a finite set of data points or uses the straight rule to estimate the properties of a population from an examination of a sample, Reichenbach admits that no method can be guaranteed to generate the correct answer any single time it is applied.²⁶ But, he argues, some methods have the important virtue of being self-corrective. It can be shown, solely by means of deductive arguments, that if they are applied repeatedly, they will get ever closer to the correct answer with the accumulation of evidence. The inductive straight rule and the principle of the simplest curve are examples of such self-corrective methods.

As Hempel notes, Reichenbach's argument is more complicated than Hempel's simplified presentation of it. The reason for this additional complexity sheds light on one of the philosophical difficulties with Reichenbach's approach. The problem is this: Many different, competing methods can be proven to be self-corrective in Reichenbach's sense. The property of asymptotic convergence is not unique to the straight rule and the principle of the simplest curve. What, then, justifies us in using these rules rather than others with the same virtue? In the case of curve fitting, the problem takes the following form. One self-corrective method—the method of linear interpolation—tells us to approximate the true function by drawing a chain of straight lines through the data points as they accumulate. But no self-respecting scientist would use that method. Instead, scientists insist on drawing smooth curves to link the data points. Why? Reichenbach suggests that it is because scientists assign a physical significance to the higher-order derivatives of the variables that define the curve. For example, if we plot distance against time, then the first-order derivative of distance with respect to time is velocity, and the second-order derivative is acceleration. Scientists want to discover laws relating those quantities to distance and time by means of continuous functions, too. So they insist on the method of the smoothest curve, rather than the method of linear interpolation, when choosing a curve to fit a set of paired values of distance and time. But even if this argument is convincing, a large number of smooth curves can still be drawn through any set of data points. Why should we prefer one curve to another that is, by whatever small degree, incompatible with it? At this point in his argument, Reichenbach concedes that there is no further inductive vindication for our preference. Inductive simplicity can take us only so far. Once we have narrowed our choice to the family of smooth curves, all of which converge asymptotically, any further preference among them is solely a matter of descriptive simplicity and convenience. In the long run, any member of the family will converge on the true function, so it does not matter which one we choose.

Hempel acknowledges that Reichenbach's argument does show something. For example, it shows that some ways of connecting data points are irrational because they lack the property of convergence. But, as Reichenbach concedes, the field of convergent methods is very wide. Reichenbach's appeal to the long run as the answer to this embarrassing variety carries little conviction, since he is unable to specify how quickly (or slowly) the convergence will take place. How satisfying is it as a philosophical solution to the problem of inductive inference to be told that a method is rational because it (along with thousands of other rival methods) will eventually lead us to the truth if only we persist for long enough (and if there is a truth to be discovered)? Most of the decisions we have to make using inductive reasoning concern the very short run. As Lord Keynes remarked, "in the long run we are all dead."

■ POPPER'S FALSIFIABILITY ANALYSIS

Popper's approach to the simplicity problem differs radically from that taken by the other philosophers discussed by Hempel. Since Popper denies the very possibility of inductive confirmation and insists that all theories, even the ones we currently accept, are almost certainly false, he can hardly claim that simplicity should be preferred because simple theories are more likely to be true. Instead, Popper links simplicity with falsifiability. Simple theories should be preferred, he tells us, because they are more falsifiable and thus easier to test. Popper also associates simplicity with empirical content, claiming that simple theories are more falsifiable because they have greater content. Inductivists and Bayesians defend the principle of simplicity by arguing that simpler theories have higher prior probabilities, that they are more likely to be true because of their simplicity.²⁷ Popper takes the opposite position, arguing that simpler theories have greater content and thus are more likely to be false. We can crudely summarize Popper's position regarding scientific theories as follows:

more falsifiable = simpler = greater empirical content
= more improbable.

There are many fascinating and controversial aspects to Popper's network of claims concerning falsifiability, simplicity, and content. Some of these are touched on in Salmon's article, discussed earlier in this commentary. Here, we shall focus on two of Popper's claims: the link between simplicity and falsifiability, and the link between simplicity and content.

In order to link simplicity with falsifiability, we must first characterize each and then show that the one tracks the other. For curves specified by equations, Popper proposes that simplicity decreases with the number of freely adjustable parameters, the degree of the equation, and the order of the derivatives (if any) that occur within it. For example, the second-order equation of a circle is simpler than the second order equation of an ellipse because it has one less parameter (the radius of the circle replaces the major and minor axes of the ellipse). Falsifiability (testability) depends on the number of data points that would be needed to show that the theory is false. Four points that suffice to refute a circle hypothesis might still lie on an ellipse. Thus "all the planets move in circles" has a greater degree of falsifiability than "all the planets move in ellipses" because, if false, it can be falsified by a smaller number of data points. This example illustrates Popper's thesis that simpler theories have a greater degree of falsifiability.

Many objections have been raised against Popper's analysis of simplicity and falsifiability. One problem is its limitation to simple polynomial functions such as those specifying straight lines, circles, and similar curves.

What about functions such as $y = \cos x$, and $y = \log(1 - x)$? Scientists would judge these to be relatively simple. Yet their power series expansions contain an infinite number of terms involving ever higher powers of x . Thus, Popper's theory would judge them all as being equally, infinitely, complex. Hempel's discussion raises two further problems. First, certain groups of hypotheses (for example, H_1 , H_2 , and H_3 from Hempel's curve-fitting example) can all be refuted by a single data point, even though some of the equations are of a higher order than others. So, it would seem, contrary to Popper, theories can differ greatly in simplicity and yet have the same degree of falsifiability. Second, when two hypotheses differ in their degree of falsifiability, the more falsifiable hypothesis is not always simpler. Consider, for example, "All asteroids and comets move in circles" (K) and "All planets move in ellipses" (L). Because it takes fewer data points to refute a circle hypothesis than it does to refute an ellipse hypothesis, K is more falsifiable than L even though it is more complex. Of course, Popper might reply by insisting that L is more falsifiable than K because it has greater content.

The content of a scientific theory is notoriously hard to define. The natural idea is to regard a theory's content as a function of the number of the empirical consequences derivable from it: the larger the set of consequences, the greater the theory's content. But most, if not all, scientific theories have an infinite number of consequences. Thus we are faced with the daunting problem of trying to decide which infinite sets are "bigger" than others. Content comparison is only straightforward when one theory entails another. For, if H entails J , then J 's consequence set is a subset of H 's consequence set. And so long as H and J are not logically equivalent, the content of H must be greater than the content of J . But, as Hempel notes, logically stronger theories are not necessarily simpler. For example, the conjunction ($H \& J$) entails J , but when H and J are unrelated theories, ($H \& J$) is not simpler than J . Greater content does not imply greater simplicity. Moreover, in the curve-fitting problem, none of the rival hypotheses entails any of the others. Nonetheless, some are simpler than others, even though, intuitively, their content is the same, since each of them specifies an infinite set of pairs of values for the two variables u and v .

THE PROBABILITY OF HYPOTHESES

In the final section of his paper, Hempel considers whether it might be possible to construct a system of inductive logic in which the credibility of a hypothesis is a precise number between 0 and 1 satisfying the axioms of probability. Since the credibility of a hypothesis, H , depends not just on a narrow set of observational evidence, E , but on the other theories, statements, and laws that we accept, Hempel writes the credibility of H as

$c(H, K)$, which is to be read as the credibility of H relative to the set K of all the statements that scientists accept at a given time (including E). Once we are given credibilities in the form of probabilities, it is easy to calculate further probabilities of disjunctions and conjunctions using the probability calculus. The philosophical challenge is to explain where the probabilities come from in the first place and what determines their values. Hempel mentions one influential attempt to do this, namely Rudolf Carnap's system (actually, infinitely many different systems) of logical probability (otherwise known as *inductive probability*, or *degree of confirmation*).

As Hempel recognizes, Carnap's work is highly formal, based as it is on an idealized model language. All the primitive predicates (of which there can be only finitely many) and individual constants (the terms that name individual objects, events, or, as Carnap prefers, spatiotemporal positions) must be completely specified at the outset and, for a significant class of these systems, the numerical values that the credibility function assigns to hypotheses depend on the number of primitive predicates in the language. Thus, if we wish to apply any of these Carnapian systems of inductive logic to science, either we must hold the language constant, refusing to allow any new predicates to be introduced, or we must recognize that the confirmation that, say, a hypothesis in physics receives from a particular body of evidence will change whenever new predicates are introduced, even if they are in completely different fields such as sociology or botany. Neither option is attractive, and in any case, the very idea that we could completely enumerate all the predicates and individual constants of a language adequate to express all the theories in present-day science is quite unrealistic.²⁸

There is another reason for being skeptical about the usefulness of Carnap's formal systems of inductive logic for understanding the confirmation of scientific theories. As Hempel points out, the confirmation of theories depends on a number of factors, such as simplicity and diversity of evidence, that are extremely difficult to analyze precisely. Thus, it is very unlikely that they could ever be incorporated into an algorithm that would determine how the numerical credibility values would change in response to them. The Bayesian approach (explored in chapter 5) circumvents this problem by abandoning Carnap's attempt to assign inductive probabilities to hypotheses on logical, a priori grounds. Instead, Bayesians take the relevant probabilities as given, empirically, in the form of the degrees of belief that scientists have in their theories. Thus, factors such as simplicity and diversity of evidence exert their influence psychologically, and Bayesian confirmation theory takes it from there. This obviates the need for the kind of precise, logical analysis that has proven so elusive.

4.5 | Snyder on Evidence, Explanation, and Novel Prediction

In her article, “Is Evidence Historical?” Laura Snyder rejects both explanationism and predictionism on the grounds that both imply the historical thesis of evidence, which, she argues, is false.

THE HISTORICAL THESIS OF EVIDENCE

The historical thesis of evidence asserts that if theory *T* entails a true statement *E*, then whether that statement confirms the theory depends on when the statement is known relative to when the theory that entails it is proposed. Explanationism entails the historical thesis because it insists that *E* can confirm *T* only if *E* is known before *T* is proposed. The view more commonly held by hypothetico-deductivists, predictionism, is equally committed to the historical thesis, since it asserts that *E* confirms *T* only if *E* is a novel prediction. Thus, predictionism holds that confirming evidence has to be “new” whereas explanationism holds that confirming evidence has to be “old.”

THREE ARGUMENTS FOR PREDICTIONISM

Snyder identifies three distinct arguments for the predictionist thesis that only new evidence can confirm a scientific theory.

■ THE NO-COINCIDENCE ARGUMENT

As Snyder discusses, predictionists and accommodationists (those who admit that old evidence can confirm but insist that novel predictions always have greater confirming power) often appeal to a form of inference to the best explanation in support of their doctrine. They point out that, when *E* is a novel prediction, there are two hypotheses that might explain why theory *T* correctly predicts *E*: the truth hypothesis and the coincidence hypothesis. Either *T* is true (and that is why *E*, along with all the other entailments of *T*, is true), or *T* is false and it is simply a coincidence that *E* happens to be true. The truth hypothesis is a much better explanation of the fact that *T* entails something true than is the coincidence hypothesis. Indeed, the coincidence hypothesis seems like no explanation at all. So when *E* is a novel prediction, *E* confirms *T* to a significant degree. But when the evidence that *T* was contrived to explain is old, a third and more likely explanation of why *T* entails that evidence enters the picture. That explanation, the design hypothesis, asserts that *T* was deliberately constructed to explain *E*. Indeed, in this case, the coincidence hypothesis drops out as a candidate, leaving the contest between the design hypothesis and the truth hypothesis. The design hypothesis is a much better expla-

Alternative explanations of the fact that a theory, *T*, entails the true statement *E*

Prediction of new evidence	<i>T</i> is true	Coincidence
Explanation of old evidence	<i>T</i> is true	Design

Table 2

nation than the truth hypothesis. So the predictionist concludes, when *E* is old evidence it lends no (or, as some accommodationists would concede, very little) support to *T*. The alternative hypotheses in the two cases are summarized in table 2.

The no-coincidence argument has several controversial aspects. One is the assumed connection between explanation and confirmation that underlies this sort of inference to the best explanation. The operative principle—that one proposition, *P*, confirms another, *Q*, only if *Q* is the best explanation for *P*—is not very plausible, in part, because confirmation comes in degrees but being the best explanation does not. Moreover, of the available explanations for *P*, *Q*, while the best, might be very poor. Under those circumstances it would seem unreasonable to regard *P* as confirming *Q* to any significant degree. These objections suggest what is perhaps a more plausible version of the explanation-confirmation principle: the degree to which *P* confirms *Q* depends on the strength of the explanation of *P* given by *Q*. The stronger the explanation of *P* by *Q*, the greater the support that *P* lends to *Q*.

Another controversial aspect of the no-coincidence argument is the judgment about which hypothesis explains best in each of the two cases. Suppose that explanation requires either the logical derivability of the explanandum (what is to be explained) from the explanans (what does the explaining) or that the explanans confer on the explanandum a reasonably high probability. The explanandum is the conjunction of two statements: that *T* entails *E*, and that *E* is true. The first conjunct of the explanandum, that *T* entails *E*, follows deductively from all of the hypotheses under consideration. For if *T* entails *E*, then this is a necessary truth and is thus entailed by any hypothesis whatever. The second conjunct of the explanandum, the fact that *E* is true, is entailed by the truth hypothesis (together with the necessary truth that *T* entails *E*). Thus, whether *E* is a novel prediction or old evidence, the hypothesis that *T* is true would seem to be an excellent explanation in either case. So why should we agree with the predictionist that, when *E* is old evidence that *T* is introduced to explain, the design hypothesis is superior to the truth hypothesis? On the face of it, the design hypothesis provides little or no explanation of the second conjunct of the explanandum, namely, that *E* is true. Granted, the

truth hypothesis beats out the coincidence hypothesis when *E* is a novel prediction. But then why does the truth hypothesis not similarly beat out the design hypothesis when *E* is old evidence?

■ THE FALSIFICATION ARGUMENT

Although Popper denies that there is any such thing as inductive confirmation, his principle of falsifiability is often invoked by inductivists. In the present debate between predictionism and explanationism, the relevant principle would be roughly as follows: if *T* entails *E*, then *E* can confirm (or “corroborate”) *T* only if *E* could potentially falsify *T*. (Paul Horwich invokes a similar principle in his solution to the raven paradox discussed in chapter 5.) When *E* is already known to be true, it is assumed that *E* cannot falsify *T*. Therefore, novel predictions always confirm and explanations of old evidence can never do so.

As presented by Snyder, the falsification argument hinges on the temporal novelty of the evidence: if *E* is already known, then it cannot serve to falsify *T*. But a number of philosophers of science, notably John Worrall, have argued that what really counts is not temporal novelty but use-novelty.²⁹ For example, if *E* was explicitly incorporated into the theory *T*, then it would not be use-novel and thus could not serve, even potentially, to falsify *T*. But *E* might have been known for centuries before *T* was proposed and yet still be able potentially to falsify *T* just as long as it was not used in the construction of *T*. For in that case, Worrall argues, *T* would be genuinely at risk from *E*. Worrall seems to have in mind *risk* in the sense of epistemic possibility, not logical possibility. From a logical point of view, either *T* entails *E*, or *T* entails something that is inconsistent with *E* (or *T* entails neither *E* nor anything that is inconsistent with *E*). Whichever entailment relation holds (or fails to hold), it does so of logical necessity, independently of what anyone believes or desires. But from an epistemic point of view, things are quite different. We want an explanation for some already known result *E*. We propose a theory *T* to explain it. Since we did not build *E* into *T* or use *E* to generate *T*, we do not know whether in fact *E* is derivable from *T*. For all we know, before we try to derive *E* from *T*, the theory might yield something that is inconsistent with *E* and thus fail the test. If we find that *T* does in fact entail *E*, then *T* has passed the test and *E* now supports *T*.

■ THE POSITIVE-RELEVANCE ARGUMENT

The positive-relevance view of evidence is the one favored by Bayesians. It is often referred to as the *relevance criterion of confirmation* or the *incremental concept of confirmation*. (See the readings and commentary in chapter 5 for an extended discussion of this approach to confirmation and evidence.) According to the positive-relevance view, *E* is evidence for the theory *T* if and only if *E* raises the probability of *T*. When a scientist uses *T* to make a novel prediction *E*, we do not yet know whether *E* is true.

Thus, the probability of E is less than 1. Assume that T entails E , so that the probability of E given T is 1: $P(E/T) = 1$. In that case, as Snyder shows, it follows immediately from Bayes's theorem that the probability of T given E is greater than the probability of T without E : $P(T/E) > P(T)$. (Again, see chapter 5 for much more on this and other applications of Bayes's theorem.) But when E is old evidence, E is already known to be true. So some philosophers have argued, in that case the probability of E is 1, and Bayes's theorem entails that E cannot change the probability of T . Hence, no matter how much old evidence T can explain, none of it can confirm T . This last claim has come to be known as the *problem of old evidence*. It should be noted that most Bayesians reject this claim and have offered various solutions to the problem of old evidence. (See chapter 5 for details.) The positive-relevance argument for predictionism is highly controversial.

TWO ARGUMENTS FOR EXPLANATIONISM

Snyder cites a recent article by Stephen Brush, "Prediction and Theory Evaluation: The Case of Light Bending," as a good example of the explanationist thesis that novel predictions are powerless to confirm theories and that only the explanation of previously known and well-established results carries confirmational weight.³⁰ In fact, Brush (and other explanationists, such as Gingerich and Worrall) does not go quite this far. Brush's position is that, while explanations of old evidence are always confirmationally superior to the prediction of new results, he does not deny that novel predictions can confirm theories to some degree. Brush rests his case on two arguments: one appeals to the history of science, the other to a notion of reliable evidence.

■ THE HISTORY-OF-SCIENCE ARGUMENT

Many science textbooks and writings by philosophers of science assume that scientists regard novel predictions as much stronger evidence for a theory than the explanation of facts already known. Wesley Salmon, for example, in his article "Rationality and Objectivity in Science or Tom Kuhn Meets Tom Bayes," (reprinted in chapter 5 below) cites the Poisson bright spot as a striking and unexpected prediction made by the wave theory of light. The story as told by Salmon, Kuhn, Giere, and others runs as follows:³¹ In 1819, Augustin Fresnel, a supporter of the wave theory of light, won the prize offered by the French Academy of Sciences for the best essay on diffraction. At that time, most scientists in France (and elsewhere) accepted the corpuscular theory of light that had been developed by Isaac Newton. Very few physicists thought that the wave theory of light could possibly be true. Of the distinguished panel of scientists that judged Fresnel's entry, three (Laplace, Biot, and Poisson) were committed corpuscularians; only one, Arago, favored the wave theory. During his ex-

amination of Fresnel's paper, Poisson deduced from Fresnel's wave theory of light the apparently absurd consequence that if a small circular disk obstructs the light coming from a pinhole, there should be a bright spot in the center of the disk's shadow and the spot should be just as bright as if the disk were not there. Arago performed an experiment to test this "absurd" consequence and, to everyone's amazement, found the bright spot, just as Fresnel's theory predicted. Fresnel was immediately awarded the prize and, almost overnight, the wave theory was accepted by the scientific community, mainly on the strength of this stunning confirmation of Fresnel's theory.

Unfortunately for the predictionists, John Worrall has shown that Arago's experimental verification of the Poisson bright spot had nothing like the dramatic effect claimed for it by most philosophers of science.¹² There was no sudden conversion to the wave theory, and the commission that judged Fresnel's paper was far more impressed by Fresnel's ability to explain, accurately and precisely, the known diffraction patterns cast by straight edges and slits, than it was by the prediction of the bright spot. The usual story about the dramatic confirming effect of the bright spot seems to be largely a myth.

Similarly mythical, according to Stephen Brush, is the story of Edington's verification, in 1919, of Einstein's novel prediction that light from distant stars will be deflected by the intense gravitational field of the sun. At the time, physicists (as opposed to the general public) were just as impressed, if not more impressed, by the ability of Einstein's general theory of relativity to predict, accurately, the rate of precession of the perihelion of Mercury. The problem of Mercury's orbit had been recognized for decades, whereas gravitational light bending was new and unexpected. But few, if any, scientists gave light bending greater confirmational significance simply because it was new. Most found Einstein's solution to the Mercury problem an achievement that was at least as, if not more, impressive than the prediction of light bending. In the view of most scientists, the power of evidence to confirm a theory has nothing to do with when the evidence was first acquired.

Thus, the history-of-science argument runs as follows: Contrary to the claims made by predictionists, scientists often accord greater weight to the explanation of known phenomena than to the prediction of new ones. This is borne out, even in those cases (such as the Poisson bright spot and gravitational light bending) to which predictionists appeal for support. Since philosophical claims about science have to square with the considered judgments of scientists in a reasonably wide variety of cases, explanationism is more plausible than predictionism.

■ THE RELIABLE-EVIDENCE ARGUMENT

Brush's appeal to the history of science as support for explanationism has a number of problems. History, that is *real* history not the simplified ide-

alizations of textbook writers, is messy. As Brush readily admits, scientists were not unanimous in the light bending case: some did think that light bending was stronger evidence for Einstein's theory than the explanation of Mercury's orbit, and Brush, a scrupulous historian, quotes some of them in his article. Of course, Brush attempts to explain why they judged the matter incorrectly or why, in some cases, they did not literally mean what they wrote. But the fact remains, then as now, that some scientists endorse predictionism.

Another way in which history is messy concerns the "other things being equal" qualification that is implicit in the rival theses of predictionism and explanationism. Both camps recognize that, even if novel predictions are intrinsically superior to explanations as evidence, in an actual case a particular explanation may confirm more strongly than a particular novel prediction and a particular novel prediction, however startling and unexpected, may not confirm a theory to any significant extent. The reason for the latter is readily appreciated from a Bayesian perspective in which confirmation depends not only on the unexpectedness of the evidence predicted but also on the prior probability of the theory making the prediction. Thus, for example, in the case cited by Owen Gingerich,³³ Velikovsky made predictions about the high surface temperature of Venus and radio emissions from Jupiter that scientists at the time regarded as highly improbable, only to have these predictions verified a few years later. Much to Velikovsky's dismay these predictive successes did nothing to increase the scientific acceptability of his theory. To Bayesians this is not surprising: the prior probability of Velikovsky's theory was so low that no amount of evidence, not even the success of startling novel predictions, could raise that probability to any significant degree. Thus, the Velikovsky case is not a counterexample to predictionism because other things were not equal.

Given the problems with appealing to history for support of explanationism, what Brush needs is a philosophical argument showing why explanations should count more strongly than novel predictions, other things being equal. In his attempt to offer such an argument, he invokes the notion of reliable evidence, which is sketched in Snyder's article. The phrase *reliable evidence* is not, perhaps, well chosen, since it suggests *probability of truth*, whereas Brush's brief remarks make clear that what he really means is something more like *confirming power*. In other words, as Brush sees it, the issue is not whether the evidence is known for sure to be genuine but whether the evidence, once known, has the power to confirm the theory.

Brush begins by noting that, prior to Einstein's general theory of relativity, not only was Mercury's orbit already well known, but, more importantly, no theory had been able to explain it. This failure did not lead to the rejection of these unsuccessful theories, since, at that time, scientists had nothing better to put in their place; Mercury's orbit was a classic example of a nonrefuting anomaly. But when Einstein solved the Mercury

problem, the unsuccessful theories were rejected, and, as a consequence, Einstein's theory was substantially confirmed. Understood in this way, Brush is appealing to inductive confirmation by the elimination of alternatives. Contrast the Mercury problem with the prediction of light bending. Gravitational light bending was an entirely new phenomenon. Thus, at the time of Einstein's successful prediction whether the possible rival theories to Einstein's could also explain it was not known. (The only available rival theory that had been systematically explored was Newton's gravitational theory, and Newton's theory did indeed give the wrong answer.) So the novel prediction of light bending, though successful, had relatively little confirming power when compared with the solution of the Mercury problem. Only some years later, when rival theories had tried and failed to explain light bending, could light bending count as strong evidence for Einstein's theory. Snyder refers to this principle—that a successful prediction cannot be good ("reliable") evidence for a theory until other theories have also had a chance at explaining it—as Brush's *equal opportunity requirement*. On this interpretation of Brush's argument, it is not the explanation of previously known phenomena per se that counts strongly in favor of a theory but the theory's succeeding where its rivals have failed.³⁴

WHY THE HISTORICAL THESIS IS FALSE

In the section of her article entitled, "The Concept of Evidence," Snyder undertakes to refute both predictionism and explanationism by showing that the historical thesis of evidence is false. Here, in schematic outline, is her argument.

- 1 As used in science, the concept of evidence is impersonal. If *E* confirms (is evidence for) hypothesis *H*, then *E* is a reason for anyone to believe *H*, regardless of who that person is and what beliefs she holds.
- 2 Therefore, any adequate philosophical theory about the nature of scientific evidence must be compatible with the impersonal concept of evidence.
- 3 Theories of evidence are either objective or nonobjective.
- 4 Objective theories of evidence deny that confirmation and evidence depend on anyone's beliefs or knowledge at any time.
- 5 Thus, all objective theories of evidence are compatible with the impersonal concept of evidence, and all entail that the historical thesis is false.
- 6 Nonobjective theories of evidence assert that confirmation and evidence do depend in specified ways on what particular individuals (or groups of individuals) know or believe at a particular time.

- 7 Only some nonobjective theories of evidence are compatible with the impersonal concept of evidence, but all of these entail that the historical thesis is false.
- 8 Therefore, the historical thesis is false, and so, too, are predictionism and explanationism.

Snyder's argument is complicated, in part because she draws a fine distinction between the concept of evidence as used by scientists being impersonal and philosophical theories about the nature of scientific evidence being objective. Snyder argues that, at bottom, *E* is evidence for some hypothesis *H* (or, equivalently, *E* confirms *H*) if and only if *E* gives a reason for thinking that *H* is true. What makes the scientific concept of evidence impersonal is that the question of whether *E* is evidence for *H* is completely independent of any particular person's beliefs. If *E* is evidence for *H*, then anyone who believes *E* has a reason for believing *H*. It does not matter whether a particular person does not believe *E* or holds other beliefs that may count against *H* or fails to appreciate the evidential connection between *E* and *H* or believes *H* for some other reason. Evidence is evidence, regardless of who considers it or what beliefs that person holds.³⁵

Many (but not all) of the theories of evidence and confirmation offered by philosophers of science are objective because they regard any subjective factor (such as the beliefs of a person or group of people) as utterly irrelevant to the confirmation relation between *E* and *H*. Since beliefs are held to be irrelevant to confirmation, it follows trivially that the time at which beliefs are acquired must also be irrelevant. So all objective theories of evidence entail that the historical thesis is false.

Perhaps the best known nonobjective theory of evidence is the subjective Bayesian view (sometimes called the *personalist* theory). On this view, confirmation is relativized to each individual so that, strictly speaking, there is no such thing as *the* confirmation that *E* gives to *H*. According to subjective Bayesians, there is only the confirmation of *H* on *E* for a particular person, and its value can vary from person to person. Snyder rejects the subjective Bayesian view as unacceptable because it violates the impersonality condition. (See chapter 5 for an extended discussion and criticism of Bayesian approaches to confirmation.) But there are other nonobjective theories of evidence that are compatible with the impersonal character of evidence in science. Snyder singles out for attention the *expert-relative concept of evidence* (see page 473):

E is evidence for *H* at time *t* only if the "experts" know some body of information *B* at *t* and, given *B*, *E* is a reason to believe *H*.

Although the expert-relative concept relativizes confirmation to the beliefs of some people ("experts") at a given time and thus is nonobjective, it is impersonal because, if the confirmation relation holds between E and H at a given time, then E is a reason to believe H at that time regardless of who one is. In particular, the expert-relative concept does not require that anyone (not even the "experts") know or believe that E is true at any given time. Thus, the expert-relative concept is inconsistent with the historical thesis (since the historical thesis says that the time at which E is known or believed is crucially relevant to whether E confirms H .)

In our reconstruction of Snyder's argument, premise number 7 says that *all* nonobjective theories of evidence that are compatible with the impersonal concept of evidence entail that the historical thesis is false. But Snyder has not shown that this is true merely by examining one such theory, the expert-relative theory. In principle, there are other possible nonobjective theories of impersonal evidence that entail that the historical thesis is true. Here is one such theory:

The predictionist "expert-relative" concept of evidence: E is evidence for H at time t only if the "experts" know some body of information B but do not know E at t and, given B , E is a reason to believe H .

Admittedly, this predictionist version of the expert-relative view sounds a little odd, but it is not totally implausible. Consider, for example, the Poisson bright spot episode discussed earlier. It is possible that J. N. Delisle had observed the bright spot at the center of a circular shadow before 1819.³⁶ But from a predictionist, "expert-relative" point of view, this would be irrelevant because the observation, would not generally have been known and would certainly not have been known by the "experts" whose beliefs determine what constitutes evidence and confirmation at a particular time. Moreover, as Snyder herself argues in the final section of her paper, we should not require that E be known for it to count as evidence for some hypothesis H . To insist otherwise, she argues, would be to conflate the distinction between E being evidence for H and E being used as evidence for H .

4.6 | Achinstein's Rejection of Predictionism and Explanationism

The title of Peter Achinstein's article asks, "Explanation v. Prediction: Which Carries More Weight?" His answer is that it all depends on the evidence in question. Like Snyder, Achinstein rejects the contention of the historical thesis that whether E confirms H *always* depends on when

	H	E
Drug case	Drug <i>D</i> relieves symptoms <i>S</i> in approximately 95 percent of cases.	In a group of 1,000 people with symptoms <i>S</i> taking drug <i>D</i> , 950 persons had relief of <i>S</i> ; in a control group of 1,000 <i>S</i> -sufferers taking, not <i>D</i> , but a placebo, none had symptoms <i>S</i> relieved.
Lottery case	John won the lottery.	In last week's lottery, 1,000 tickets were sold of which John owned 999 at the time of the selection of the winner; this was a fair lottery in which one ticket was selected at random.
Crow case	Female crows are black.	Males crows are black.

Table 3

E is known relative to when *H* is proposed. But unlike Snyder, Achinstein thinks that whether *E* is evidence for *H* does *sometimes* depend on historical facts about how *E* was collected. Nonetheless, even when evidence is historical in this sense, Achinstein argues that it makes no difference whether *E* is an explanation of old evidence or a novel prediction. More generally, when extra information is relevant to evidence and confirmation in particular cases, sometimes the extra information is historical; sometimes it is not. What kind of information is relevant in particular cases is an empirical matter. Thus, Achinstein is sharply critical of confirmation theories such as Carnap's, which attempt to answer the question, "How strongly does *E* confirm *H*" solely on a priori grounds.

SELECTION PROCEDURES

Achinstein's argument for his verdict that history sometimes is and sometimes is not relevant to evidence rests on the consideration of different cases. In the section of his article on selection procedures, he offers three such cases: the drug case, the lottery case, and the crow case. Their essential features are summarized in table 3.

Achinstein argues that in the drug case whether *E* is evidence for *H* (and certainly, how strong that evidence is) depends on the selection procedures used to choose the experimental and control groups. Without this historical information, we simply cannot tell how strongly *E* supports *H*

or whether *E* supports *H* at all. Contrast this with the lottery case, in which *E* is strong evidence for *H* no matter what method was used to find out that *E* is true. Because *E* says that the lottery was fair, that the winning ticket was picked at random, and that John held all but one of the 1,000 tickets in the lottery, everything inductively relevant to *H* is already contained in *E*. Finally, in the crow case, like the drug case, whether *E* is evidence for *H* (and if so, how strong that evidence is) depends on extra, empirical information. But unlike the drug case, the extra information needed is ornithological, not historical. Using the terminology introduced by Achinstein, in the lottery case, *E* is empirically complete with respect to *H*. In the crow case, *E* is empirically incomplete with respect to *H*, but no historical facts are needed to settle claims about evidence and confirmation. So the lottery and crow cases are both counterexamples to the historical thesis of evidence.

PREDICTION VERSUS EXPLANATION

If, as Snyder has argued, the historical thesis is false, then both predictionism and explanationism (in their pure, strong forms defined earlier) must be false. But even if the predictionist is mistaken in asserting that explanations can never be evidence and the explanationist in asserting that novel predictions can never be evidence, the interesting question remains whether, in cases in which history is relevant to confirmation, predictions or explanations provide the stronger evidence.

Obviously, in cases that violate the historical thesis (and hence in which temporal considerations are irrelevant to confirmation), *E* confirms (or fails to confirm) *H* regardless of whether evidence *E* is old or new. Achinstein reinforces this point with a coin-tossing example that, like the lottery case, is empirically complete. The more interesting cases are those in which the evidence *E* is historically incomplete. For an example of such a case, Achinstein returns to the drug case. The information contained in *E* can be construed either as *E*₁—a prediction about what will occur in the next trial—or as *E*₂—a report about a trial that has already taken place. And just to sharpen the contrast, Achinstein supposes that *H* (the drug hypothesis) was deliberately devised in order to explain *E*₂. So which is stronger evidence for *H*, *E*₁ or *E*₂?

Achinstein thinks that the correct answer to this question is “neither.” For if (as Achinstein argued in the previous section) the confirming power of evidence *E* depends on the selection procedure used to generate it, then when our knowledge of *E* was first acquired makes no difference whatever. If *E*₁ is strong evidence for *H* because the trial will involve randomization over a diverse population, then *E*₂ must be equally strong evidence for *H* if the same selection procedure was used in a trial that already took place. The explanation-versus-prediction distinction is irrelevant.

Achinstein does not say what, in general, makes for a good selection procedure relative to a hypothesis H . But a theory implicit in his remarks is made explicit in an important article by Deborah Mayo that Achinstein cites in a footnote.³⁷ Mayo's proposal is that E is good evidence for H if and only if E is the outcome of a severe test of H . A test of H is severe if there is a very low probability that the test would have yielded a passing result such as E if H were false. A "passing result" in this context means a result that accords with H at least as well as E does. When H is a statistical hypothesis, as in Achinstein's drug case, a small range of outcomes (including E) would qualify as passing the test of H . When H is a deterministic hypothesis that entails E , then no other outcome could accord with H at least as well.

But in either case, the crucial question in judging the severity of the test, and hence the power of E to confirm H , is "How likely is it that we would have obtained a result that is at least as good as E if H were false?" In the deterministic case, when H entails E , this boils down to requiring that $P(E/\sim H)$ —the probability of E given that H is false—be very low. As Mayo points out, predictionists often assume, falsely, that this requirement will be violated whenever E has been used to generate H (and thus E is not use-novel or novel in any other of the senses we have distinguished).

Mayo gives a simple example to illustrate why the predictionists are mistaken in thinking that severe tests entail use-novelty. Suppose that the fender of a car has been dented while it was parked in a garage. The car's owner examines carefully the shape and size of the dent and uses this information, E , to infer H , the likely make of the car that caused the dent. If it is practically impossible for the dent to have its distinctive features unless it was created by a collision with a specific type of car tailfin, then E is good evidence for H , even though E was used to generate H .

BRUSH REDUX

The word *redux* means "restored," but in the section "Brush Redux" Achinstein neither reinstates Brush's explanationist position nor defends it. (This is hardly surprising given Achinstein's preceding attack on explanationism and predictionism.) Instead, Achinstein criticizes two of the theses (the ones that Achinstein numbers 1 and 3) that Brush appeals to when trying to show why explanations of known phenomena confirm theories more strongly than do novel predictions. Both theses concern the presumed desirability of considering competitors to H when assessing a selection procedure for testing H . Thesis (1) says that a test of H is inferior if it fails to consider competitors to H , and thesis (3) says that scientists have more time to consider competitors when explaining known phenomena than when making novel predictions. Achinstein regards his drug case as a counterexample to thesis (1), since at least one of the selection procedures for generating E in that example (SP 1) results in a strong test of

the drug hypothesis H , even though the procedure makes no explicit mention of competitors. And in any case, Achinstein continues, even if thesis (1) were true, there is no reason to think that thesis (3) must be true. For if what determines the strength of a test is the character of the selection procedure, then a test will be strong or weak regardless of whether the evidence generated by it is old or new. And this will hold even if strong tests have to consider competitors to the hypothesis being tested.

Achinstein's arguments against theses (1) and (3) are not entirely convincing. It is true that the selection procedure (SP 1) that he alludes to in criticizing thesis (1) does not mention competitors to H explicitly. (SP 1) says that the experimental and control groups should be chosen "arbitrarily" (483)—presumably, Achinstein means at random—from a diverse population of subjects. But the whole point of randomization and diversity is to diminish the chances that the results obtained could be due to any cause other than drug D that is being tested. The beauty of randomization is that we do not have to formulate explicitly the competing hypotheses to H in order to be pretty sure that we would not have obtained the result E unless H were true. So while the consideration of competitors does not figure in (SP 1) explicitly, it is an important part of the implicit rationale for adopting it. Moreover, when we turn our attention from simple statistical generalizations (as in the drug case) to hypotheses in physics and other sciences, it is hard to see how we could we begin to estimate $P(E/\sim H)$ for a test of H unless we had some idea of the likely ways in which H might be false. And that, surely, will often require the explicit consideration of the competitors to H .

THOMSON VERSUS HERTZ

Mindful of the highly artificial and simplified character of the cases on which so much of his argument depends, Achinstein concludes by giving a real-life example from the history of physics. When nineteenth-century physicists began applying a high electrical potential to gases at low pressures, it became evident that something was being emitted in straight lines from the negative electrode (the cathode). What was the nature of these cathode rays (as they were soon named)? There were two competing theories: either they were streams of negatively charged particles, or they were some kind of ether wave (electromagnetic radiation). In 1883, Heinrich Hertz (1857–94) reasoned as follows:³⁸ If they were charged particles, then they should be deflected both by a magnetic field and by an electrostatic field. Since the cathode rays did bend in response to a magnet but apparently did not bend in response to an electrostatic field, Hertz concluded that the rays must not be charged particles but some peculiar form of ether wave. (Peculiar because, unlike light rays, they were deflected by a magnetic field.³⁹) In 1897, J. J. Thomson (1856–1940) challenged Hertz's

conclusion on the grounds that Hertz's experiment was flawed. Because the pressure in Hertz's cathode tube had not been low enough, the gas molecules had become ionized by the radiation streaming from the electrodes. These ionized gas molecules effectively short-circuited the plates to which an electrostatic charge had been applied, thus destroying the electric field.⁴⁰ When the tube was evacuated more thoroughly, Thomson found the deflection that Hertz had been unable to detect.

Achinstein makes two points about this episode. First, Hertz was wrong to take his negative result *E* (the failure to detect any electrostatic deflection) as strong evidence for *H* (the hypothesis that cathode rays are not electrically charged). In fact, as Thomson later showed, Hertz's procedure for testing *H* was inadequate, and hence *E* is not evidence for *H*. Second, and more important in Achinstein's view, is that, as far as we can tell, Hertz did not propose hypothesis *H* until after he had performed his experiment. But no one (certainly not Thomson in 1897) suggested that, had Hertz first proposed *H* and then predicted *E* *before* his experiment, the confirmation (or, as it turns out, the lack of confirmation) that *E* confers on *H* would have been affected. The crucial flaw in Hertz's experiment, his use of a poor selection procedure, prevents *E* from supporting *H*, regardless of whether *H* was used to predict *E* in advance of the experiment or, as actually happened, *H* was introduced after the experiment to explain its outcome.

4.7 | Summary

The only thing concerning induction about which all philosophers agree is that inductive arguments are not deductively valid. After that, the controversy begins. Although we could simply define as inductive any argument in which the premises do not logically entail the conclusion, most philosophers follow the custom of identifying as inductive a select subclass of invalid arguments. This subclass can be picked out either by enumerating particular forms of argument (such as statistical syllogism and arguments from analogy) and designating them as inductive or by relying on the normative judgment that inductive arguments are those that confer on their conclusions a reasonably high degree of probability. As Peter Lipton explains, both strategies confront difficulties. The first strategy has to solve the problem of description, the second—the problem of justification.

The problem of description is the problem of articulating the general principles involved when we reason inductively. Although several simple forms of inductive argument have been identified, they are only a small part of the entire field of inductive reasoning, which includes both inferences to new hypotheses (in the context of discovery) and arguments used to support the claim that a hypothesis is probable, well confirmed, or ac-

ceptable (in the context of justification). Lipton examines four simple models of inductive inference and finds each wanting as a description of our inductive practice.

Even if the problem of description could be solved, the problem of justification would remain. For all inductive arguments, from the simplest to the most complex, face the problem posed by David Hume. Hume argued that it is impossible to show that inductive arguments lead from true premises to true conclusions with a reasonably high probability. A crucial step in Hume's reasoning is his insistence that to justify our future confidence in induction by appealing to its past success would be to beg the question: such an attempt at justification would be circular, since it, too, relies on inductive reasoning.

As explained in the commentary on Lipton, the challenge to induction raised by Hume may not be as threatening as it appears at first sight. Even if Hume's skeptical argument is sound, it does not establish that no inductive inference is justified or even that no inductive argument can justify our belief that induction is reliable. At best what Hume may have proved is that we cannot show an inductive skeptic, someone who is unwilling to admit that any nondeductive argument justifies its conclusion, that induction is reliable. When one reflects on the similar difficulty that would confront anyone trying to justify deductive reasoning to a deductive skeptic, it becomes less clear that Hume's problem is a serious challenge to the rationality of induction.

The many attempts to solve the problem of induction fall into four broad categories. Some philosophers claim that induction is rational and justified simply because of what *rational* and *justified* mean in the English language. Others reject this ordinary-language approach and give inductive arguments to support induction, arguing that Hume was wrong to condemn them as circular. John Stuart Mill (in the nineteenth century) and Max Black (in the twentieth) are advocates of this sort of inductive answer to Hume's skepticism. Still others (notably Reichenbach and Salmon) agree with Hume that inductive arguments in favor of induction are viciously circular and concede that inductive arguments cannot be shown to be justified. Nonetheless, these philosophers claim that our practice of using induction can be vindicated pragmatically by showing that it is a reliable tool for reaching the truth. The overall strategy of the vindication approach is to argue that, while there is no guarantee that induction will succeed, it can be shown (using only deductively valid arguments) that repeated applications of inductive reasoning will eventually take us closer to the truth, if indeed there is a truth of the matter to be reached.

Finally, there are Popper and his followers who think that induction can be neither justified nor vindicated. But they believe that science is rational despite this because it aims, not at the confirmation of hypotheses, but at their falsification. Science progresses by proposing ever bolder hypotheses (which Popper often calls *conjectures*) and then trying to refute

them by comparing their deductive consequences with the results of observation and experiment (in the guise of what Popper calls *basic statements*). Theories that have survived severe tests and have not yet been refuted are *corroborated*. *Corroboration* is not the same as *confirmation*. Corroborated theories are neither probable nor confirmed, since, according to Popper, inductive probability and confirmation are myths. Rather, to say that a theory has been corroborated is to give a historical report on the severity of the trials that it has weathered.

Popper's falsificationism has been attacked from many angles. Some critics have charged that, because basic statements are simply accepted by convention in Popper's system, every decision either to reject or not to reject a hypothesis in light of those statements must ultimately be conventional and arbitrary. Others have challenged Popper's claim that following the method of bold conjectures and refutations will lead science in the direction of ever greater truth-content or verisimilitude. Wesley Salmon attacks Popper on the grounds that, without inductive confirmation, Popper (and his defender, John Watkins) cannot solve the problem of rational prediction, that is, the problem of explaining why we should use corroborated theories (as opposed to other as-yet-unfalsified theories) for making inferences about the future. Salmon focuses specifically on those inferences that are part of practical decision making. A key premise in Salmon's criticism of Popper is the assumption that, if corroboration is solely about a theory's past performance, then it cannot serve as a rational guide to the future unless one indulges in inductive reasoning. For, Salmon asks, how else can the past track record of a theory (which is all that judgments of corroboration amount to, according to Popper and Watkins) be relevant to our rational expectations about the future without our making an inference—an inductive inference—from the one to the other? Thus, Salmon concludes that we cannot make sense of science as a rational enterprise if we follow Popper. Scientists must make use of inferences that go beyond deduction, and Popper's falsificationism cannot be the whole story about scientific method.

Like Salmon and most other philosophers of science, Carl Hempel is convinced that Popper has failed to provide a complete and adequate model of scientific reasoning. Although Hempel acknowledges that falsification plays an important role in science, he contends that it is only part of the picture. The rest of the picture is confirmation, and Hempel undertakes a survey of the various criteria that scientists use in deciding when a theory is sufficiently well confirmed to be accepted as probable or true (at least until further notice). Hempel is especially interested in whether a convincing case can be made for these criteria as indicators of truth. He examines, in turn, quantity and variety of evidence, novelty of predictions, support from other theories, and simplicity. His discussion of these factors is inevitably rather sketchy, and he deliberately avoids taking a definite position in the debates concerning them.

Several points emerge from Hempel's survey of criteria of confirmability and acceptability. First, given the diversity of these criteria, it is not surprising that Hempel's discussion of them is rather piecemeal. He neither seeks nor offers a single, unified rationale for them all. In fact, at the end of his article, Hempel expresses serious doubts about whether these factors could be incorporated into a confirmation theory such as Carnap's in which there is a unique, *a priori*, degree of inductive probability that any piece of evidence confers on any theory. (In chapter 5, we examine the attempt by Bayesians to fit Hempel's criteria into a unified framework.)

Second, Hempel reveals that a genuine puzzle arises from the predictionist claim that novel predictions must always confirm a theory better than explanations of previously known facts, for predictionism entails that a result that strongly confirms a theory in one context (in which the result is new) might not confirm the theory to any appreciable degree in another context (in which the result is already known). The predictionist tries to dispel the air of paradox here by deploying the so-called no-coincidence or no-surprise argument: for a theory correctly to predict a novel result if the theory were not true would be very surprising, but for a theory designed to explain a previously known fact to get it right would be no surprise at all. This argument is examined at length, along with other arguments for predictionism, by Laura Snyder. Hempel briefly rehearses a possible rejoinder to the no-coincidence argument for predictionism that appeals to the notion of simplicity: that a theory yields the result it was designed to explain may not be surprising, but that the theory yielding the result in question is a simple one *is* surprising. This response raises the question, which Hempel addresses later in his article, of whether there is any connection between theoretical simplicity and probability of truth.

Third, the notion of theoretical support from "above" appears to be uncontroversial: whatever confirms a theory also confirms anything that can be derived from that theory, including other theories. That consequences inherit confirmation from any confirmed theory that entails them is presupposed whenever we gather evidence for a theory in order to have rational confidence in its predictions. But as Hempel has argued elsewhere, accepting top-down confirmation (in the form of the special consequence condition) means that we have to place some restrictions on bottom-up confirmation. One simple principle that might be thought to govern bottom-up confirmation is the converse consequence condition, according to which evidence that confirms a given theory also confirms any stronger theory. But the two conditions together entail that any evidence confirms any theory, which is unacceptable. Thus, in his other writings, Hempel has rejected the converse consequence condition and, with it, the naive version of the hypothetico-deductive (H-D) model. Contrary to the naive version of the H-D model, hypotheses are not automatically confirmed by their true deductive consequences. Rather, Hempel

has argued, only the consequences of a theory that are instances of that theory have the power to confirm it.

Fourth, Hempel shows how difficult it is to characterize simplicity in precise terms and then to connect it with truth or probable truth. Rejecting the views of Mach and others who regarded simplicity as solely a matter of descriptive economy, Hempel explores in detail the contrasting theories of Reichenbach and Popper. Reichenbach, the inductivist, defends simplicity as a guide to truth by the same form of vindication argument that he offers in defense of induction. In either case, whether following the inductive straight rule (what Lipton calls "More of the Same") or choosing the simplest theory in light of the available evidence, Reichenbach claims that repeated applications of this procedure will allow us to converge on the correct answer in the long run. But, Reichenbach's critics object, convergence is a property shared by many different procedures, not just the inductive straight rule and the simplicity principle. And, besides, without some indication of how rapidly the convergence will occur (or, in other words, how long the "long run" is likely to be) Reichenbach's argument seems to be too weak to vindicate either the inductive straight rule or our preference for simplicity in theories. Popper, the falsificationist, denies that simpler theories are more probable but argues that they are to be preferred because they are more falsifiable and have greater empirical content. As Hempel explains, neither of these claims can withstand close examination.

Laura Snyder attacks both predictionism and explanationism (the view that only the explanation of previously known results can confirm a theory) by attacking the historical thesis of evidence to which both positions are committed. The historical thesis asserts that when evidence becomes known is relevant to confirmation. Snyder criticizes the historical thesis on the grounds that it conflicts with the impersonal nature of evidence as working scientists conceive of it. To say that evidence is impersonal means that, if some result is evidence for a theory, then the result is a reason to believe that the theory is true regardless of who considers the matter or what that person's beliefs are. Snyder argues that any adequate philosophical theory of scientific evidence must reflect this impersonality. Since objective theories of evidence deny that human beliefs and knowledge have anything to do with what makes something evidence for a theory, they entail that the historical thesis is false. Some nonobjective theories of evidence, such as the personalist, subjective variety of Bayesianism, unabashedly violate the impersonality condition, and Snyder rejects them for that reason. But others, specifically the expert-relative concept of evidence, respect the impersonality condition. The expert-relative concept of evidence, though nonobjective (because it refers to the background beliefs of experts as an essential part of its characterization of what evidence is), is also inconsistent with the historical thesis. Thus, Snyder concludes that the historical thesis and all the positions that entail it (such as predictionism and explanationism) are false.

One possible flaw with Snyder's argument is that she has not exhaustively examined all the possible impersonal, nonobjective theories of evidence to see whether one or two might be consistent with the historical thesis. Such a survey is necessary in the absence of any general philosophical argument that the impersonality of evidence, by its very nature, is inconsistent with the historical thesis. In the discussion of Snyder's article in this commentary, a possible counterexample to her argument is suggested.

There is a wide measure of agreement between Achinstein and Snyder concerning the falsity of the historical thesis and the irrelevance of the explanation-versus-prediction distinction to questions of evidence. But Achinstein argues that historical information can sometimes affect the strength of evidence, especially when the results being used as evidence for a theory were obtained by performing an experiment. In cases involving drug trials, for example, the evidential value of the results crucially depends on how the experimental and control groups were selected. That one selection procedure rather than another was employed is a historical fact about the experiment. So regardless of whether the experiment was performed before or after the formulation of a theory—and hence regardless of whether the report of the outcome of the experiment is old or new evidence—the report's status as evidence depends on historical facts. Achinstein also gives an example from the history of physics to support his contention that selection procedures, not the explanation-versus-prediction distinction, are what counts in determining whether an experimental outcome is or is not evidence for a theory.

■ | Notes

1. C. D. Broad, "The Philosophy of Francis Bacon," in *Ethics and the History of Philosophy* (New York: Humanities Press, 1952), 117–43.
2. Gilbert Harman, "The Inference to the Best Explanation," *Philosophical Review* 74 (1965): 88–95. Richard Fumerton takes the opposite tack, arguing that arguments to the best explanation are good only insofar as they can be reduced to inductive (and deductive) reasoning. See Richard A. Fumerton, "Induction and Reasoning to the Best Explanation," *Philosophy of Science* 47 (1980): 589–600.
3. Hume's own version of the argument is different from ours, mainly because Hume divided arguments into the demonstrative and the experimental. (Hume himself never used the term *induction*.) Demonstrative arguments, for Hume, are deductively valid and all their premises are necessary truths or what Hume called *relations of ideas*. Experimental arguments have among their premises at least one empirical statement or what Hume called a *matter of fact or existence*. Thus, Hume would classify all inductive arguments as well as some deductively valid arguments as experimental. Since the principle of the uniformity of nature is empirical (a matter of fact or existence), it cannot be the conclusion of any demonstrative

argument. (No empirical statement can be validly entailed by a set of premises unless at least one of those premises is also empirical.) So the uniformity principle can only be the conclusion of an experimental argument. Hume then argued that any such argument will beg the question if we attempt to use it to justify our belief in the uniformity principle.

4. See, for example, James Van Cleve, "Reliability, Justification, and the Problem of Induction," in P. French, T. Uehling, Jr., and H. Wettstein, eds., *Midwest Studies in Philosophy*, vol. 9 (Minneapolis: University of Minnesota Press, 1984), 555–67.

5. The following analysis owes much to Susan Haack, "The Justification of Induction," *Mind* 85 (1976): 112–19; James Van Cleve, "Reliability, Justification, and the Problem of Induction"; and D. H. Mellor, "The Warrant of Induction," *Matters of Metaphysics* (Cambridge: University of Cambridge Press, 1991), 254–68.

6. For good introductory surveys of attempts to solve Hume's problem of justifying induction, see Wesley C. Salmon, *The Foundations of Scientific Inference* (Pittsburgh: University of Pittsburgh Press, 1967); Wesley C. Salmon, "Unfinished Business: The Problem of Induction," *Philosophical Studies* 33 (1978): 1–19; Brian Skyrms, *Choice and Chance: An Introduction to Inductive Logic*, 3d ed. (Belmont, Calif.: Wadsworth, 1986); and Richard G. Swinburne, ed., *The Justification of Induction* (Oxford: Oxford University Press, 1974).

7. Hans Reichenbach introduced the terms *context of discovery* and *context of justification* in the first chapter of his book, *Experience and Prediction* (Chicago: University of Chicago Press, 1938). For Reichenbach, the distinction was logical rather than temporal. Whether before or after a theory is formulated, anything describing psychological events belongs to the context of discovery, anything evaluating the logical relations among propositions belongs to the context of justification. Thus, for Reichenbach, there cannot be a "logic of discovery" in the sense of a "logic of the context of discovery," since the latter context is defined as being purely a psychological matter. Nonetheless, Reichenbach thought it possible to give a rational reconstruction of the inferences involved in the discovery of scientific theories. See Martin Curd, "The Logic of Discovery: An Analysis of Three Approaches," in *Scientific Discovery, Logic, and Rationality*, ed. Thomas Nickles (Dordrecht, Netherlands: D. Reidel, 1980), 201–19.

8. For a discussion of Goodman's grue problem and the new riddle of induction, see "Goodman's Gruesome New Riddle of Induction" in the commentary on chapter 3.

9. For the raven paradox and an attempted solution, see "The Raven Paradox" in the commentary on chapter 5.

10. Let *H* be the hypothesis that all crows are black and all swans are purple, let *E* be the statement that this bird is black, and let *O* be the observation report that this bird is a crow. Although *E* does not follow from *O* all by itself, *E* is entailed by the conjunction of *O* with *H*. Therefore, according to the H-D model, given *O*, *E* confirms *H*. But if a hypothesis is confirmed, so too are the consequences derivable from it. *H* entails that all swans are purple. Thus, our observation of a black crow confirms the prediction of purple swans. For more on this problem,

see "Theoretical Support" in the discussion of Hempel's article later in this commentary.

11. The other three methods are the joint method of agreement and difference, the method of residues, and the method of concomitant variations. (Despite its name, the joint method of agreement and difference has nothing to do with the method of difference.) There is a thorough treatment of Mill's methods and their limitations in Brian Skyrms, *Choice and Chance: An Introduction to Inductive Logic*. The account in this commentary derives from Lilly-Marlene Russow and Martin Curd, *Principles of Reasoning* (New York: St. Martin's Press, 1989), ch. 10.

12. Later in his book, *Inference to the Best Explanation* (from which the reading in this chapter is excerpted), Lipton changes and expands Mill's methods in an attempt to overcome some of these limitations.

13. See, for example, Karl Popper, *Conjectures and Refutations* (New York: Harper and Row, 1963), 47–48, 190–93.

14. For a good introduction to the debate over the possibility of a logic of discovery, see the lengthy introductory essay by Thomas Nickles in *Scientific Discovery, Logic, and Rationality*, 1–59, and Curd, "The Logic of Discovery: An Analysis of Three Approaches."

15. To forestall a possible misunderstanding, it should be recognized that Popper is not using the phrase *basic statement* to mean a statement that is completely justified or certain or one that can be conclusively verified by immediate observation. Rather, for Popper, basic statements are singular existential assertions about particular regions of space and time. Popper gives as an example: "There is a raven in space-time region *k*." Popper calls such statements *basic*, not because they have any special epistemological status, but because they are used to test and falsify theories. See Karl Popper, *The Logic of Scientific Discovery* (New York: Basic Books, 1959), ch. 5.

16. Popper, *The Logic of Scientific Discovery*, 111.

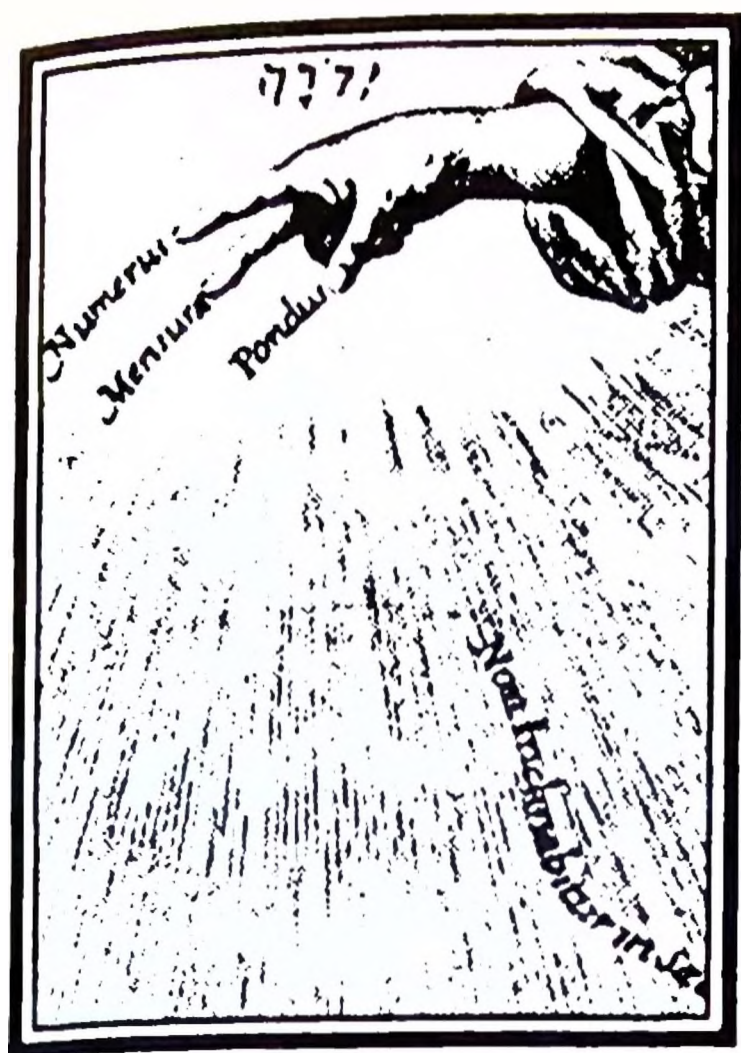
17. These and other searching criticisms of Popper are made with clarity and vigor in W. H. Newton-Smith, *The Rationality of Science* (Boston, Mass.: Routledge and Kegan Paul, 1981), ch. 3. See also Adolf Grünbaum, "Is the Method of Bold Conjectures and Attempted Refutations *Justifiably* the Method of Science?" *British Journal for the Philosophy of Science* 27 (1976): 105–36. Both Newton-Smith and Grünbaum dispute Popper's claim that his falsificationist methodology will lead to theories of greater truth-content or what Popper calls *verisimilitude*.

18. In a more recent publication, Watkins has conceded the force of Salmon's objections. See John Watkins, *Science and Scepticism* (Princeton, N.J.: Princeton University Press, 1984), ch. 9.

19. Elie Zahar, "Why Did Einstein's Programme Supersede Lorentz's? (I) and (II)," *British Journal for the Philosophy of Science* 24 (1973): 103. Alan Musgrave and others have criticized this notion of novelty as being too subjective, since it depends on the intentions of the person who designed the theory. See pp. 12–15 of Alan Musgrave, "Logical versus Historical Theories of Confirmation," *British Journal for the Philosophy of Science* 25 (1974): 1–23.

20. John Worrall, "Scientific Discovery and Theory-Confirmation," in *Change and Progress in Modern Science*, ed. Joseph C. Pitt (Dordrecht, Netherlands: D. Reidel, 1985), 301–31. The distinction between design-novelty and use-novelty is also discussed in Deborah G. Mayo, "Novel Evidence and Severe Tests," *Philosophy of Science* 58 (1991): 523–52.
21. See Carl G. Hempel, "Studies in the Logic of Confirmation," *Mind* 54 (1945): 1–26, 97–121; reprinted in C. G. Hempel, *Aspects of Scientific Explanation* (New York: Macmillan, 1965), 3–46.
22. It also entails the equivalence condition, an essential premise in the raven paradox. See "The Raven Paradox" in the commentary on chapter 5 for details.
23. This argument presupposes what, in the commentary on chapter 5, we call the absolute concept of confirmation. On this view, *E* confirms *H* if and only if the probability of *H* given *E* exceeds some suitably high threshold value, say, 0.9. On the relevance criterion of confirmation favored by Bayesians, the special consequence condition fails. According to the relevance criterion, *E* confirms *H* if and only if *E* raises the probability of *H*. Consider the following example, given by Richard Jeffrey. Suppose that there is a simple eight-ticket lottery. *E*: the winner is ticket 2 or 3. *H*: the winner is ticket 3 or 4. *G*: the winner is neither ticket 1 nor 2. $P(H) = 1/4$, and $P(H/E) = 1/2$. Thus, *E* confirms *H*. Also, of course, *H* entails *G*. But $P(G) = 3/4$, and $P(G/E) = 1/2$. So *E* does not confirm *G*, thus violating the special consequence condition. See Richard C. Jeffrey, "Probability and the Art of Judgment," in *Observation, Experiment, and Hypothesis in Modern Physical Science*, ed. Peter Achinstein and Owen Hannaway (Cambridge: MIT Press, 1985), 95–126.
24. More precisely, Hempel's satisfaction criterion asserts that a hypothesis *H* is confirmed by an observation report *B* if *H* is satisfied in the finite class of individuals mentioned (essentially) in the report. For further details, see Hempel, "Studies in the Logic of Confirmation."
25. Hans Reichenbach, *Experience and Prediction*, 374.
26. The straight rule is a version of the principle of more of the same. It says that when a certain fraction of *As* are *B* in a sample, infer that the same fraction of *As* are *B* in the population from which the sample was drawn.
27. See "Simplicity" in the commentary on chapter 5.
28. For this and other criticisms of Carnap's project, see Ernest Nagel, "Carnap's Theory of Induction," in *The Philosophy of Rudolf Carnap*, ed. P. A. Schilpp (LaSalle, Ill.: Open Court, 1963), 785–826.
29. See John Worrall, "Fresnel, Poisson and the White Spot: The Role of Successful Predictions in the Acceptance of Scientific Theories," in *The Uses of Experiment*, ed. David Gooding, Trevor Pinch, and Simon Schaffer (Cambridge: Cambridge University Press, 1989), 135–57, and the discussion of Zahar in Alan Musgrave, "Logical versus Historical Theories of Confirmation."
30. Stephen G. Brush, "Prediction and Theory Evaluation: The Case of Light Bending," *Science* 246 (1989): 1124–29. See also Stephen G. Brush, "Dynamics of Theory Change: The Role of Predictions," in *PSA 1994*, vol. 2, ed. D. Hull,

- M. Forbes, and R. M. Burian (East Lansing, Mich.: Philosophy of Science Association, 1994), 133–45.
31. Thomas S. Kuhn, *The Structure of Scientific Revolutions*, 2d ed. (Chicago, Ill.: University of Chicago Press, 1970), 155; Ronald N. Giere, "Testing Theoretical Hypotheses," in *Testing Scientific Theories*, *Minnesota Studies in the Philosophy of Science*, vol. 10, ed. J. Earman (Minneapolis: University of Minnesota Press, 1983), 279–84. Significantly, both accounts rely on the same source for their historical information: E. T. Whittaker, *A History of the Theories of Aether and Electricity*, vol. 1 (London: Thomas Nelson and Sons, 1910), 108.
32. See Worrall, "Fresnel, Poisson, and the White Spot."
33. Owen Gingerich, "Neptune, Velikovsky and the Name of the Game," *Scientific American* 275 (September 1996): 181–83.
34. Alan Musgrave defends a similar condition as being the best version of the historical approach to confirmation. Musgrave attributes the core idea to Lakatos. See Musgrave, "Logical versus Historical Theories of Confirmation," 15–19.
35. Contrast Snyder's characterization of scientific evidence as impersonal with the contextual analysis of evidence defended by Helen Longino in chapter 2. As noted in the commentary on chapter 2, Longino's analysis is not inconsistent with the objective concept because Longino is analyzing what we take to be evidence whereas Snyder is interested in what evidence really is, regardless of what we take it to be in a particular context.
36. Worrall cites Whittaker (1910) as the source of this twist in the bright spot saga but says that he has not been able to verify it. See Worrall (1989), 152–53.
37. Deborah G. Mayo, "Novel Evidence and Severe Tests."
38. Just three years later, in 1886, Hertz achieved international recognition for his experimental verification of the existence of radio waves predicted by Maxwell's theory of electromagnetism.
39. As Michael Faraday had discovered, a magnetic field will rotate the plane of polarization of polarized light, but it will not cause a light beam to bend. Moreover, as Achinstein points out in his book, *Particles and Waves*, Hertz also discovered in 1891 that cathode rays could penetrate thin layers of metal. These metal foils were far too thick to allow molecules or atoms to pass through them, but light was known to be able to penetrate thin films of gold. This was another result that Hertz took to be evidence against the particle hypothesis and in favor of the ether-wave hypothesis. See Peter Achinstein, *Particles and Waves* (New York: Oxford University Press, 1991), 283. For a different perspective on Hertz's experiment, see Jed Z. Buchwald, "Why Hertz Was Right about Cathode Rays," in *Scientific Practice: Theories and Stories of Doing Physics*, ed. Jed Z. Buchwald (Chicago: University of Chicago Press, 1995), 151–69.
40. Rom Harré, *Great Scientific Experiments* (Oxford: Phaidon, 1981), 176.



5 | Confirmation and Relevance: Bayesian Approaches

INTRODUCTION

Two years after his death in 1761, there appeared in the *Philosophical Transactions of the Royal Society* a paper entitled “An Essay Towards Solving a Problem in the Doctrine of Chances” by the Reverend Thomas Bayes: nonconformist preacher, fellow of the Royal Society, and mathematician.¹ From such inauspicious beginnings has arisen an astonishing variety of doctrines and theories that bear Bayes’s name, ranging from work in the foundations of probability theory and statistical inference to models of scientific rationality and theories of confirmation. There are mathematicians, statisticians, and philosophers of every stripe and persuasion—objectivists, subjectivists, personalists, tempered personalists, falsificationists, eliminative inductivists, and hypothetico-deductivists—who call themselves “Bayesians.” In this chapter we will examine some of the main Bayesian approaches to understanding the confirmation of scientific theories.

A considerable part of the attraction of the Bayesian approach is that, on the basis of a simple theorem of probability—Bayes’s theorem—and a few assumptions about rationality and degrees of belief, Bayesianism promises a unified explanation of a wide range of accepted principles and truisms of scientific methodology. These include the special evidential value we attach to surprising predictions, our preference for simple hypotheses and our aversion to ad hoc ones when rival theories confront the same evidence, our conviction that diverse sets of evidence lend stronger support to theories than do narrow ones, and the fact that not every theory is confirmed equally well by the evidence it entails. Unlike the simple version of the hypothetico-deductive (H-D) method (explored in chapter 4), which regards these and other principles as separate and unconnected, Bayesianism sees them all as flowing from the central underlying logic of

confirmation encapsulated in Bayes's equation. In addition, Bayesianism also offers to resolve various paradoxes of confirmation such as the raven paradox and Goodman's new riddle of induction. It is an offer that few modern philosophers of science have been able to refuse.

In "Rationality and Objectivity in Science or Tom Kuhn Meets Tom Bayes," Wesley C. Salmon attempts to reconcile Kuhn's historically based philosophy of science with the more traditional approach of the logical empiricists. The key to this reconciliation, Salmon argues, lies in paying proper attention to the role of prior probabilities in determining degrees of confirmation. Salmon candidly admits that some aspects of the application of Bayes's equation to confirmation, such as the calculation of the expectedness of evidence, present intractable problems. In response, Salmon advocates a modified version of Bayesianism that skirts these problems but still provides an algorithm for comparing rival theories when they are judged by the same evidence.

Clark Glymour's "Why I Am Not a Bayesian" has become something of a modern classic in the annals of Bayesian philosophy of science. Glymour systematically examines all the major problems with the Bayesian approach—including the weakness of standard arguments for the foundational assumptions of Bayesianism, the difficulty in relating the simplicity of theories to their prior probability, and the problem of old evidence—and criticizes some of the attempts that have been made to solve them. Apart from his specific criticisms, Glymour has a deeper complaint against Bayesianism, namely, its alleged inability to shed any interesting light on what makes evidence relevant to the theory it confirms. As a consequence, even when Bayesianism can be made compatible with methodological principles and truisms, Glymour charges that its account of them is superficial and unenlightening.

Paul Horwich replies to Glymour and other critics of Bayesianism in his article, "Wittgensteinian Bayesianism." He emphasizes the value of "therapeutic Bayesianism" as a philosophical cure for the paradoxes of confirmation and for understanding the relevance of simplicity to inductive support. Horwich replies to critics who attack the foundational assumptions of the Bayesian approach, arguing that a certain sort of criticism is inappropriate—he calls it "misplaced scientism"—in response to a theory aiming to solve philosophical problems rather than to give a complete and accurate description of scientific method.

■ | Notes

1. For a splendid account of Bayes's essay, see John Earman, *Bayes or Bust?* (Cambridge, Mass.: MIT Press, 1992).

WESLEY C. SALMON

Rationality and Objectivity in Science or Tom Kuhn Meets Tom Bayes

Twenty-five years ago, as of this writing, Thomas S. Kuhn published *The Structure of Scientific Revolutions*.¹ It has been an extraordinarily influential book. Coming at the height of the hegemony of *logical empiricism*—as espoused by such figures as R. B. Braithwaite, Rudolf Carnap, Herbert Feigl, Carl G. Hempel, and Hans Reichenbach—it posed a severe challenge to the logistic approach that they practiced.² It also served as an unparalleled source of inspiration to philosophers with a historical bent. For a quarter of a century there has been a deep division between the logical empiricists and those who adopt the historical approach, and Kuhn's book was undoubtedly a key document in the production and preservation of this gulf.

At a 1983 meeting of the American Philosophical Association (Eastern Division), Kuhn and Hempel—the most distinguished living advocates for their respective viewpoints—shared the platform in a symposium devoted to Hempel's philosophy.³ I had the honor to participate in this symposium. On that occasion Kuhn chose to address certain issues pertaining to the rationality of science that he and Hempel had been discussing for several years. It struck me that a bridge could be built between the differing views of Kuhn and Hempel if Bayes's theorem were invoked to explicate the concept of scientific confirmation.⁴ At the time it seemed to me that this maneuver could remove a large part of the dispute between standard logical empiricism and the historical approach to philosophy of science on this fundamental issue.

I still believe that we have the basis for a new consensus regarding the choice among scientific theories. Although such a consensus, if achieved, would not amount to total agreement on every problem, it would

FROM C. Wade Savage, ed., *Scientific Theories*, vol. 14, *Minnesota Studies in the Philosophy of Science* (Minneapolis: University of Minnesota Press, 1990), 175–204.

represent a major rapprochement on an extremely fundamental issue. The purpose of the present essay is to develop this approach more fully. As it turns out, the project is much more complex than I thought in 1983.

1 | Kuhn on Scientific Rationality

A central part of Kuhn's challenge to the logical empiricist philosophy of science concerns the nature of theory choice in science. The choice between two fundamental theories (or paradigms), he maintains, raises issues that "cannot be resolved by proof." To see how they are resolved we must talk about "techniques of persuasion," or about "argument and counter-argument in a situation in which there can be no proof." Such choices involve the exercise of the kind of judgment that cannot be rendered logically explicit and precise. Such statements, along with many others that are similar in spirit, led a number of critics to attribute to Kuhn the view that science is fundamentally irrational and lacking in objectivity.

Kuhn was astonished by this response, which he regarded as a serious misinterpretation. In his "Postscript—1969," in the second edition of *The Structure of Scientific Revolutions*, and in "Objectivity, Value Judgment, and Theory Choice"⁵ he replies to these charges. What he had intended to convey was the claim that the decision by the community of trained scientists *constitutes* the best criterion of objectivity and rationality we can have. In order better to understand the nature of such objective and rational methods we need to look in more detail at the considerations that are actually brought to bear by scientists when they endeavor to make comparative evaluations of competing theories.

For purposes of illustration, Kuhn offers a (nonexhaustive) list of characteristics of good scientific theories that are, he claims:

individually important and collectively sufficiently varied to indicate what is at stake. . . . These five characteristics—accuracy, consistency, scope, simplicity, and fruitfulness—are all standard criteria for evaluating the adequacy of a theory. . . . Together with others of much the same sort, they provide the shared basis for theory choice.⁶

Two sorts of problems arise when one attempts to use them.

Individually the criteria are imprecise: individuals may legitimately differ about their applicability to concrete cases. In addition, when deployed together, they repeatedly prove to conflict with one another; accuracy may, for example, dictate the choice of one theory, scope the choice of its competitor.⁷

For reasons of these sorts—and others as well—individual scientists may, at a given moment, differ regarding a particular choice of theories. In the course of time, however, the interactions among individual members of the community of scientists produce a consensus for the group. Individual choices inevitably depend upon idiosyncratic and subjective factors; only the outcome of the group activity can be considered objective and fully rational.

One of Kuhn's major claims seems to be that observation and experiment, in conjunction with hypothetico-deductive reasoning, do not adequately account for the choice of scientific theories. This has led some philosophers to believe that theory choice is not rational. Kuhn, in contrast, has tried to locate the additional factors that are involved. These additional factors constitute a crucial aspect of scientific rationality.

2 | Bayes's Theorem

The first step in coming to grips with the problem of evaluating and choosing scientific hypotheses or theories⁸ is the recognition of the inadequacy of the traditional hypothetico-deductive (H-D) schema as a characterization of the logic of science. According to this schema, we confirm a scientific hypothesis by deducing from it, in conjunction with suitable initial conditions and auxiliary hypotheses, an observational prediction that turns out to be true. The H-D method has a number of well-known shortcomings. (1) It does not take account of alternative hypotheses that might be invoked to explain the same prediction. (2) It makes no reference to the initial plausibility of the hypothesis being evaluated. (3) It cannot accommodate cases, such as the testing of statistical hypotheses, in which the observed outcome is not deducible from the hypothesis (in conjunction with the pertinent initial conditions and auxiliary hypotheses), but only rendered more or less probable.

In view of these and other considerations, many logical empiricists agreed with Kuhn regarding the inadequacy of hypothetico-deductive confirmation. A number—including Carnap and Reichenbach—appealed to Bayes's theorem, which may be written in the following form:*

$$P(T|E.B) = \frac{P(T|B)P(E|B.T)}{P(T|B)P(E|B.T) + P(\sim T|B)P(E|B.\sim T)} \quad 1$$

* Salmon uses a dot to stand for "and" and a vertical slash to indicate a conditional probability. Thus, $P(E|B.T)$ should be read as "the probability of E given B and T ." Other authors write this as $P(E/B\&T)$.

Let “T” stand for the theory or hypothesis being tested, “B” for our background information, and “E” for some new evidence we have just acquired. Then the expression on the left-hand side of the equation represents the probability of our hypothesis on the basis of the background information and the new evidence. This is known as the *posterior probability*. The right-hand side of the equation contains four probability expressions. Two of these, $P(T|B)$ and $P(\sim T|B)$, are called *prior probabilities*; they represent the probability, on the basis of background information alone, without taking account of the new evidence E, that our hypothesis is true or false respectively. Obviously the two prior probabilities must add up to one; if the value of one of them is known, the value of the other can be inferred immediately. The remaining two probabilities, $P(E|T.B)$ and $P(E|\sim T.B)$, are known as *likelihoods*; they are, respectively, the probability that the new evidence would occur if our hypothesis were true and the probability that it would occur if our hypothesis were false. The two likelihoods, in contrast to the two prior probabilities, must be established independently; the value of one does not automatically determine the value of the other. To calculate the posterior probability of our hypothesis, then, we need three separate probability values to plug into the right-hand side of Bayes’s theorem—a prior probability and two likelihoods.

Before attempting to resolve any important issues concerning the nature of scientific reasoning, let us look at a simple and noncontroversial application of Bayes’s theorem. Consider a factory that produces can openers at the rate of 6,000 per day. This factory has two machines, a new one that produces 5,000 can openers per day and an old one that produces 1,000 per day. Among the can openers produced by the new machine 1 percent are defective; among those produced by the old machine 3 percent are defective. We pick one can opener at random from today’s production and find it defective. What is the probability that it was produced by the new machine?

We can get the answer to this question via Bayes’s theorem. If we let “B” stand for the class of can openers produced in this factory today, “T” for the class of can openers produced by the new machine, and “E” for a can opener that is defective, then the probability we seek is the posterior probability $P(T|B.E)$ —the probability that a defective can opener from today’s production was produced by the new machine. The values of the prior probabilities and likelihoods have been given, namely:

$$\begin{array}{ll} P(T|B) = 5/6 & P(\sim T|B) = 1/6 \\ P(E|T.B) = 1/100 & P(E|\sim T.B) = 3/100. \end{array}$$

Plugging these values into equation (1) immediately yields $P(T|B.E) = 5/6$. Notice that the old machine has a greater probability of producing a defective can opener than does the new, but the probability that a defective

can opener was produced by the new machine is greater than that it was produced by the old one. This results, obviously, from the fact that the new machine produces so many more can openers overall than does the old one.

One way to look at this example is to consider the hypothesis T that a given can opener was produced by the new machine. This is a causal hypothesis. Our background information B is simply that the can opener is part of today's production at this factory. On the basis of this prior information, we can evaluate the prior probability of T ; it is $\frac{5}{6}$. Now we add to our knowledge about this can opener the information E that it is defective. This knowledge is relevant to the hypothesis that it was produced by the new machine; the posterior probability is $\frac{5}{8}$. Although one does not *need* to appeal to Bayes's theorem to establish this result,⁹ the highly artificial example shows clearly just how Bayes's theorem can be used to ascertain the posterior probability of a simple causal hypothesis.

When we come to more realistic scientific cases, it is not so easy to see how to apply Bayes's theorem; the prior probabilities may seem particularly difficult. I believe that, in fact, they reflect the plausibility arguments scientists often bring to bear in their deliberations about scientific hypotheses. I shall discuss this issue in [section] 4; indeed, in subsequent sections we shall have to take a close look at all of the probabilities that enter into Bayes's theorem.

In this section I have been concerned to present Bayes's theorem and to make a few preliminary remarks about its application to the problem of evaluating scientific hypotheses. In the next section I shall try to spell out the connections between Bayes's theorem and Kuhn's views on the nature of theory choice. Before moving on to that discussion, however, I want to present two other useful forms in which Bayes's theorem can be given. In the first place, because of the theorem on total probability:

$$P(E|B) = P(T|B) P(E|T.B) + P(\sim T|B) P(E|\sim T.B), \quad 2$$

equation (1) can obviously be rewritten as:

$$P(T|E.B) = \frac{P(T|B) \times P(E|T.B)}{P(E|B)}. \quad 3$$

In the second place, equation (1) can be generalized to handle several alternative hypotheses, instead of just one hypothesis and its negation, as follows:

$$P(T_i|B.E) = \frac{P(T_i|B) \times P(E|T_i.B)}{\sum_{i=1}^k [P(T_i|B) \times P(E|T_i.B)]}, \quad 4$$

where $T_1 - T_k$ are mutually exclusive and exhaustive alternative hypotheses and $1 \leq i \leq k$.

Strictly speaking, (4) is the form that is needed for realistic historical examples—such as the corpuscular (T_1) and wave (T_2) theories of light in the nineteenth century. In that case, although we could construe T_1 and T_2 as mutually exclusive, we could not legitimately consider them exhaustive, for we cannot be sure that one or the other is true. Therefore, we would have to introduce T_3 —what Abner Shimony has called the *catchall hypothesis*—which says that T_1 and T_2 are both false. $T_1 - T_3$ thus constitute a mutually exclusive and exhaustive set of hypotheses. This is the sort of situation that obtains when scientists are attempting to choose a correct hypothesis from among two or more serious candidates.

3 | Kuhn and Bayes

For purposes of discussion, Kuhn is willing to admit that “each scientist chooses between competing theories by deploying some Bayesian algorithm which permits him to compute a value for $P(T|E)$, i.e., for the probability of the theory T on the evidence E available both to him and the other members of his professional group at a particular period of time.”¹⁰ He then formulates the crucial issue in terms of the question of whether there is one unique algorithm used by all rational scientists, yielding a unique value for P , or whether different scientists, though fully rational, may use different algorithms yielding different values of P . I want to suggest a third possibility to account for the phenomena of theory choice—namely, that many different scientists might use the same algorithm, but nevertheless arrive at different values of P .

When one speaks of a Bayesian algorithm, the first thought that comes to mind is Bayes’s theorem itself, as embodied in any of the equations (1), (3) or (4). We have, for instance:

$$P(T|E.B) = \frac{P(T|B) \times P(E|B.T)}{P(E|B)}, \quad 3$$

which constitutes an algorithm in the most straightforward sense of the term. Let us call $P(E|B)$ the *expectedness* of the evidence. Given values for the prior probability, likelihood, and expectedness, the value of the posterior probability can be computed by trivial arithmetical operations.¹¹

If we propose to use equation (3) as an algorithm, the obvious question is how to get values for the expressions on the right-hand side. Several answers are possible in principle, depending on what interpretation of the probability concept is espoused. If one adopts a Carnapian approach to inductive logic and confirmation theory, all of the probabilities that appear

in Bayes's theorem can be derived a priori from the structure of the descriptive language and the definition of degree of confirmation.* Since it is extremely difficult to see how any genuine scientific case could be handled by means of the highly restricted apparatus available within that approach, not many philosophers are tempted to follow this line. Moreover, even if a rich descriptive language were available, it is not philosophically tempting to suppose that the probabilities associated with serious scientific theories are a priori semantic truths.

Two major alternatives remain. First, one might maintain that the probabilities on the right-hand side of (3)—especially the prior probability $P(T|B)$ —are objective and empirical. I have attempted to defend the view that they refer, at bottom, to the frequencies with which various kinds of hypotheses or theories have been found successful.¹² Clearly, enormous difficulties are involved in working out that alternative; I shall return to the issue below. In the meantime, let us consider the other—far more popular—alternative.

The remaining alternative approach involves the use of personal probabilities. Personal probabilities are subjective in character; they represent subjective degrees of conviction on the part of the individual who has them, provided that they fulfill the condition of coherence.¹³ Consider a somewhat idealized situation. Suppose that, in the presence of background knowledge B (which may include initial conditions, boundary conditions, and auxiliary hypotheses) theory T deductively entails evidence E . This is the situation to which the hypothetico-deductive method appears to be applicable. In this case, $P(E|T.B)$ must equal 1, and equation (3) reduces to:

$$P(T|E.B) = P(T|B)/P(E|B). \quad 5$$

One might then ask a particular scientist for his or her plausibility rating of theory T on background knowledge B alone, quite irrespective of

* Salmon is referring to the formal approach to inductive logic pioneered by Rudolf Carnap. The Carnapian approach starts with a well-defined artificial language. Every well-formed sentence in that language gets assigned a numerical value for its prior probability via an a priori choice of a measure function over the state descriptions expressible in the language. (A state description is a statement that describes a possible state of the world in as much detail as the language permits.) The axioms and definitions of probability theory can then be used to calculate the degree of inductive probability (confirmation or partial entailment) that any one sentence or conjunction of sentences confers on any other. As Salmon notes, it is very difficult to see how Carnap's formal treatment of inductive logic could be applied to real-life scientific theories. For a simple introduction to Carnap's approach, see Rudolf Carnap, "Statistical and Inductive Probability," in *The Structure of Scientific Thought*, ed. E. H. Madden. (Boston: Houghton Mifflin, 1960), 269–79, and Henry E. Kyburg Jr., *Probability and Inductive Logic* (London: Collier-Macmillan, 1970), ch. 5.

whether evidence E obtains or not. Likewise, the same individual may be queried regarding the degree to which evidence E is to be expected irrespective of the truth or falsity of T . According to the personalist, it should be possible—by direct questioning or by some less direct method—to elicit such *psychological* facts regarding a scientist involved in investigations concerning the theory in question. This information is sufficient to determine the degree of belief this individual should have in the theory T given the background knowledge B and the evidence E , namely, the posterior probability $P(T|E.B)$.

In the more general case, when T and B do not deductively entail E , the procedure is the same, except that the value of $P(E|T.B)$ must also be ascertained. In many contexts, where statistical significance tests can be applied, a value of the likelihood $P(E|T.B)$ can be calculated, and the personal probability will coincide with the value thus derived. In any case, whether statistical tests apply or not, there is no *new* problem in principle involved in procuring the needed degree of confidence. This reflects the standard Bayesian approach in which all of the probabilities are taken to be personal probabilities.

In any case, whether one adopts an objective or a personalistic interpretation of probability, equation (3)—or some other version of Bayes's theorem—can be taken as an algorithm for evaluating scientific hypotheses or theories. Individual scientists, using the same algorithm, may arrive at different evaluations of the same hypothesis because they plug in different values for the probabilities. If the probabilities are construed as objective, different individuals may well have different estimates of these objective values. If the probabilities are construed as personal, different individuals may well have different subjective assessments of them. Bayes's theorem provides a mechanical algorithm, but the judgments of individual scientists are involved in procuring the values that are to be fed into it. This is a general feature of algorithms; they are not responsible for the data they are given.

4 | Prior Probabilities

In [section] 2 I remarked that the prior probabilities in Bayes's theorem can best be seen as embodying the kinds of plausibility judgments that scientists regularly make regarding the hypotheses with which they are concerned. Einstein, who was clearly aware of this consideration, contrasted two points of view from which a theory can be criticized or evaluated:

The first point of view is obvious: the theory must not contradict empirical facts. . . . [it] is concerned with the confirmation of the theoretical foundation

by the available empirical facts. The second point of view is not concerned with the relation of the material of observation but with the premises of the theory itself, with what may briefly but vaguely be characterized as the "naturalness" or "logical simplicity" of the premises. . . . The second point of view may briefly be characterized as concerning itself with the "inner perfection" of a theory, whereas the first point of view refers to the "external confirmation."¹⁴

Einstein's second point of view is the sort of thing I have in mind in referring to plausibility arguments or judgments concerning prior probabilities.

Plausibility considerations are pervasive in the sciences; they play a significant—indeed, *indispensable*—role. This fact provides the initial reason for appealing to Bayes's theorem as an aid to understanding the logic of evaluating scientific hypotheses. Plausibility arguments serve to enhance or diminish the probability of a given hypothesis prior to—i.e., without reference to—the outcome of a particular observation or experiment. They are designed to answer the question, "Is this the kind of hypothesis that is likely to succeed in the scientific situation in which the scientist finds himself or herself?" On the basis of their training and experience, scientists are qualified to make such judgments.

This point can best be explained, I believe, in terms of concrete examples. Since before the time of Newton, for instance, a well-known plausibility argument for the inverse square character of gravitational forces has been around. It is natural to think of the gravitational force emanating from a particle of matter as one that spreads spherically from it in a uniform manner. In the seventeenth and eighteenth centuries all competent physical scientists believed that physical space has a three-dimensional Euclidean structure. Since the surface of a Euclidean sphere increases as the square of the radius, it is reasonable to suppose that the force of gravity is diluted in just the same way, for the farther one goes from the particle, the greater the spherical surface over which the force must be spread.

A famous Canadian study of the effects of the consumption of large doses of saccharin provides another example.¹⁵ A statistically significant association between heavy saccharin consumption and bladder cancer in a controlled experiment with rats lends considerable plausibility to the hypothesis that use of saccharin as an artificial sweetener in diet soft drinks increases the risk of bladder cancer in humans. This example, unlike the preceding one, is inherently statistical and does not have even the *prima facie* appearance of a hypothetico-deductive inference.

In order to come to a clearer understanding of the nature of prior probabilities, it will be necessary to look at them from the point of view of the personalist and that of the objectivist (frequency or propensity theorist).¹⁶ The frightening thing about pure unadulterated personalism is that nothing prevents prior probabilities (and other probabilities as well) from

being determined by all sorts of idiosyncratic and objectively irrelevant considerations. A given hypothesis might get an extremely low prior probability because the scientist considering it has a hangover, has had a recent fight with his or her lover, is in passionate disagreement with the politics of the scientist who first advanced the hypothesis, harbors deep prejudices against the ethnic group to which the originator of the hypothesis belongs, etc. What we want to demand is that the investigator make every effort to bring all of his or her *relevant* experience in evaluating hypotheses to bear on the question of whether the hypothesis under consideration is of a type likely to succeed, and to leave aside emotional irrelevancies.

It is rather easy to construct really perverse systems of belief that do not violate the coherence requirement. But we need to keep in mind the objectives of science. When we have a long series of events, such as tosses of fair or biased coins, or radioactive decays of unstable nuclei, we want our subjective degrees of conviction to match what either a frequency theorist or a propensity theorist would regard as the objective probability. Carnap was profoundly correct in his notion that inductive or logical or epistemic probabilities should be reasonable estimates of relative frequencies.

A sensible personalist, I would suggest, is someone who wants his or her personal probabilities to reflect objective fact. Betting on a sequence of tosses of a coin, a personalist wants not only to avoid Dutch books,¹⁷ but also to stand a reasonable chance of winning (or of not losing too much too fast). As I read it, the whole point of F. P. Ramsey's famous article on degrees of belief is to consider what you get if your subjective degrees of belief match the relevant frequencies.¹⁸ One of the facts recognized by the sensible personalist is that whether the coin lands heads or tails is not affected by on which side of the bed he or she got out that morning. If we grant that the personalist's aim is to do as well as possible in betting on heads and tails, it would be obviously counterproductive to allow the betting odds to be affected by such irrelevancies.

The same general sort of consideration should be brought to bear on the assignment of probabilities to hypotheses. Whether a particular scientist is dyspeptic on a given morning is irrelevant to the question of whether a physical hypothesis that is under consideration is correct or not. Much more troubling, of course, is the fact that any given scientist may be inadvertently influenced by ideological or metaphysical prejudices. It is obvious that an unconscious commitment to capitalism or racism might seriously affect theorizing in the behavioral sciences.

Similar situations may arise in the physical sciences as well; another historical example will illustrate the point. In 1800 Alessandro Volta invented the battery, thereby providing scientists with a way of producing steady electrical currents. It was not until 1820 that Hans Christian Oersted discovered the effect of an electrical current on a magnetic needle. Why was there such a delay? One reason was the previously established

fact that a static electric charge has no effect on a magnetic needle. Another reason that has been mentioned is the fact that, contrary to the expectation if there were such an effect, it aligns the needle perpendicular to the current carrying wire. As Holton and Brush remark, "But even if one has currents and compass needles available, one does not observe the effect unless the compass is placed in the right position so that the needle can respond to a force that seems to act in a direction *around* the current rather than *toward* it."¹⁹ I found it amusing when, on one occasion, a colleague set up the demonstration with the magnetic needle oriented at right angles to the wire to show why the experiment fails if one begins with the needle in that position. When the current was turned on, the needle rotated through 180 degrees; he had neglected to take account of polarity. How many times, between 1800 and 1820, had the experiment been performed without reversing the polarity? Not many. The experiment had apparently not been tried by others because of Cartesian metaphysical commitments. It was undertaken by Oersted as a result of his proclivities toward *naturphilosophie*.*

How should scientists go about evaluating the prior probabilities of hypotheses? In elaborating a view he calls *tempered personalism*—a view that goes beyond standard Bayesian personalism by placing further constraints on personal probabilities—Shimony²⁰ points out that experience shows that the hypotheses seriously advanced by serious scientists stand some chance of being successful. Science has, in fact, made considerable progress over the past four or five centuries, which constitutes strong empirical evidence that the probability of success among members of this class is nonvanishing. Likewise, hard experience has also taught us to reject claims of scientific infallibility. Thus, we have good reasons for avoiding

* By "Cartesian metaphysical commitments," Salmon is referring to the fact that, prior to the nineteenth century, the three inverse-square law forces that had been discovered—of gravity, electricity, and magnetism—were either attractions or repulsions that acted in a straight line between their sources. To most physicists it seemed inevitable that all forces must have this linear, push-and-pull character. Oersted's discovery, in 1820, that the interaction between a current-carrying wire and a magnet was rotatory and polar in nature, came as a profound shock. To observe the reversal of polarity noted by Salmon, put the wire above the compass needle and at right angles to it, with the current running from east to west. The needle will swivel through 180 degrees, clearly indicating the magnetic polarity of the force that circulates around the wire. Salmon also refers to the role of *naturphilosophie* in Oersted's discovery. *Naturphilosophie* (philosophy of nature) was an antimechanistic, holistic, and idealist system of thought associated with the German philosopher Friedrich Schelling (1775–1854). Its main features were an insistence on the fundamental unity of nature (in the form of a few basic patterns or archetypes) and the appeal to opposed forces or "polarities" to explain diversity. It motivated Oersted to search for an underlying connection between electricity and magnetism and made him receptive to the possibility of polar forces that do not conform to the mechanistic, "Cartesian" push-pull pattern.

the assignment of extreme values to the priors of the hypotheses with which we are seriously concerned. Moreover, Shimony reminds us, experience has taught that science is difficult and frustrating; consequently, we ought to assign fairly low prior probabilities to the hypotheses that have been explicitly advanced, allowing a fairly high prior for the catchall hypothesis—the hypothesis that we have not yet thought of the correct hypothesis. The history of science abounds with situations of choice among theories in which the successful candidate has not even been conceived at the time.

In *The Foundations of Scientific Inference*, I proposed that the problem of prior probabilities be approached in terms of an objective interpretation of probability, in particular, the frequency interpretation. I suggested three sorts of criteria that can be brought to bear in assessing the prior probabilities of hypotheses: formal, material, and pragmatic.

Pragmatic criteria have to do with the circumstances in which a new hypothesis originates. We have already seen an example of a pragmatic criterion in Shimony's observation that hypotheses advocated by serious scientists have nonvanishing chances of success. The opposite side of the same coin is provided by Martin Gardner, who offers an enlightening characterization of scientific cranks.²¹ Since it is doubtful that a single useful scientific suggestion has ever been originated by anyone in that category, hypotheses advanced by people of that ilk have negligible chances of being correct. I recall when L. Ron Hubbard's *Dianetics* was first published. A psychologist friend, asked what he thought of it, said, "I can't condemn this book before reading it, but after I have read it, I will." When competent scientists offer hypotheses outside of their areas of specialization, we have a right to wonder whether appreciable plausibility accrues to such suggestions. Hubbard was, incidentally, an engineer with no training in psychology.

The *formal criteria* have to do not only with matters of internal consistency of a new hypothesis, but also with relations of entailment or incompatibility of the new hypothesis with accepted laws and theories. The fact that Immanuel Velikovsky's *Worlds in Collision*²² contradicts many of the accepted basic laws of physics—e.g., the law of conservation of angular momentum—renders his 'explanations' of such biblically reported incidents as the parting of the waters of the Red Sea and the brief interruption of the rotation of the earth (the sun standing still) utterly implausible.

It should be recalled that among his five considerations for the evaluation of scientific theories—mentioned above—Kuhn includes consistency of the sort we are discussing. I take this as a powerful hint that one of the main issues Kuhn has raised about scientific theory choice involves the use of prior probabilities and plausibility judgments.

The *material criteria* have to do with the actual structure and content of the hypothesis or theory under consideration. The most obvious example is simplicity—another of Kuhn's five items. Simplicity strikes me as

singularly important, for it has often been treated by scientists and philosophers as an a priori criterion. It has been suggested, for example, that the hypothesis that quarks are fundamental constituents of matter loses plausibility as the number of different types of quarks increases, since it becomes less simple as a result.²³ It has also been advocated as a universal methodological maxim: *Search for the simplest possible hypothesis*. Only if the simpler hypotheses do not stand up under testing should one resort to more complex hypotheses.

Although simplicity has obviously been an important consideration in the physical sciences, its applicability in the social/behavioral sciences is problematic. In a recent article, "Slips of the Tongue," Michael T. Motley criticizes Freud's theory for being too simple—an oversimplification.

Further still, the categorical nature of Freud's claim that all slips have hidden meanings makes it rather unattractive. It is difficult to imagine, for example, that my six-year-old daughter's mealtime request to "help cut up my meef" was the result of repressed anxieties or anything of that kind. It seems more likely that she simply merged "meat" and "beef" into "meef." Similarly, about the only meaning one can easily read into someone's saying "roon mock" instead of "moon rock" is that the m and r got switched. Even so, how does it happen that words can merge or sounds can be switched in the course of speech production? And in the case of my "pleased to beat you" error [to a competitor for a job], might Freud have been right?²⁴

The most reasonable way to look at simplicity, I think, is to regard it as a highly relevant characteristic, but one whose applicability varies from one scientific context to another. Specialists in any given branch of science make judgments about the degree of simplicity or complexity that is appropriate to the context at hand, and they do so on the basis of extensive experience in that particular area of scientific investigation. Since there is no precise measure of simplicity as applied to scientific hypotheses and theories, scientists must use their judgment concerning the degree of simplicity a given hypothesis or theory possesses and concerning the degree of simplicity that is desirable in the given context. The kind of judgment to which I refer is not spooky; it is the kind of judgment that arises on the basis of training and experience. This experience is far too rich to be the sort of thing that can be spelled out explicitly. As Patrick Suppes²⁵ has pointed out, the assignment of prior probability by the Bayesian can be regarded as the best estimate of the chances of success of the hypothesis or theory on the basis of all relevant experience in that particular scientific domain. The personal probability represents, not an effort to contaminate science with subjective irrelevancies, but rather an attempt to facilitate the inclusion of all relevant evidence.

Simplicity is only one among many material criteria. Another closely related criterion—frequently employed in contemporary physics—is sym-

metry. Perhaps the most striking historical example is de Broglie's hypothesis regarding matter waves. Since light exhibits both particle and wave behavior, which are linked in terms of linear momentum, he suggested, why should not material particles, which obviously possess linear momentum, also have such wave characteristics as wavelength and frequency? Unbeknownst to de Broglie, experimental work by Davisson was, at that very time, providing positive evidence of wavelike behavior of electrons.

A third widely used material criterion is analogy, as illustrated by the saccharin study. The physiological analogy between rats and humans is sufficiently strong to lend considerable plausibility to the hypothesis that saccharin can cause bladder cancer in humans. I suspect that the use of arguments by analogy in science is almost always aimed at establishing prior probabilities. The formal criteria enable us to take account of the ways in which a given hypothesis fits *deductively* with what else we know. Analogy helps us to assess the degree to which a given hypothesis fits *inductively* with what else we know.

The moral I would draw concerning prior probabilities is that they can be understood as our best estimates of the frequencies with which certain kinds of hypotheses succeed. These estimates are rough and inexact; some philosophers might prefer to think of them in terms of intervals. If, however, one wants to construe them as personal probabilities, there is no harm in it, as long as we attribute to the subject who has them the aim of bringing to bear all his or her experience that is relevant to the success or failure of hypotheses similar to that being considered. The personalist and the frequentist need not be in any serious disagreement over the construal of prior probabilities.²⁶

One point is apt to be immediately troublesome. If we are to use Bayes's theorem to compute values of posterior probabilities, it would appear that we must be prepared to furnish numerical values for the prior probabilities. Unfortunately, it seems preposterous to suppose that plausibility arguments of the kind we have considered could yield exact numerical values. The usual answer is that, because of a phenomenon known as "washing out of the priors" or "swamping of the priors," even very crude estimates of the prior probabilities will suffice for the kinds of scientific judgments we are concerned to make. Obviously, however, this sort of convergence depends upon agreement regarding the likelihoods.

5 | The Expectedness

The term " $P(E|B)$ " occurring in the denominator of equation (3) is called the *expectedness* because it is the opposite of surprisingness. The smaller the value of $P(E|B)$, the more surprising E is; the larger the value of $P(E|B)$, the less surprising, and hence, the more expected E is. Since

the expectedness occurs in the denominator, a smaller value tends to increase the value of the fraction. This conforms to a widely held intuition that the more surprising the predictions a theory can make, the greater is their evidential value when they come true.

A classic example of a surprising prediction that came true is the Poisson bright spot. If we ask someone who is completely naive about theories of light how probable it is that a bright spot appears in the center of the shadow of a brightly illuminated circular object (ball or disk), we would certainly anticipate the response that it is very improbable indeed. There is a good inductive basis for this answer. In our everyday lives we have all observed many shadows of opaque objects, and they do not contain bright spots at their centers. Once, when I demonstrated the Poisson bright spot to an introductory class, one student carefully scrutinized the ball bearing that cast the shadow because he strongly suspected that it had a hole through it.

Another striking example, to my mind, is the Cavendish torsion-balance experiment. If we ask someone who is totally ignorant of Newton's theory of universal gravitation how strongly they expect to find a force of attraction between a lead ball and a pith ball in a laboratory, I should think the answer, again, would be that it is very unlikely. There is, in this example as well, a sound inductive basis for the response. We are all familiar with the gravitational attraction of ordinary-size objects to the earth, but we do not have everyday experience of an attraction between two such relatively small (electrically neutral and unmagnetized) objects as those Cavendish used to perform his experiment. Newton's theory predicts, of course, that there will be a gravitational attraction between any two material objects. The trick was to figure out how to measure it.

As the foregoing two examples show, there is a possible basis for assigning a low value to the expectedness; it was made plausible by assuming that the subject was completely naive concerning the relevant physical theory. The trouble with this approach is that a person who wants to use Bayes's theorem—in the form of equation (3), say—cannot be totally innocent of the theory T that is to be evaluated, since the other terms in the equation refer explicitly to T . Consequently, we have to recognize the relationship between $P(E|B)$ and the prior probabilities and likelihoods that appear on the right-hand side in the theorem on total probability:

$$P(E|B) = P(T|B) P(E|T.B) + P(\sim T|B) P(E|\sim T.B). \quad 2$$

Suppose that the prior probability of T is not negligible and that T , in conjunction with suitable initial conditions, entails E . Under these circumstances E cannot be totally surprising; the expectedness cannot be vanishingly small. Moreover, to evaluate the expectedness of E we must also consider its probability if T is false. By focusing on the expectedness, we cannot really avoid dealing with likelihoods.

There is a further difficulty. Suppose, for example, that the wave theory of light is true. It is surely *true enough* in the context of the Poisson bright spot experiment. If we want to evaluate $P(E|B)$ we must include in B the initial conditions of the experiment—the circular object illuminated by a bright light in such a way that the shadow falls upon a screen. Given the truth of the wave theory, the *objective probability* of the bright spot is one, for whenever those initial conditions are realized, the bright spot appears. It makes no difference whether we know that the wave theory is true, or believe it, or reject it, or have ever thought of it. Under the conditions specified in B the bright spot invariably occurs. Interpreted either as a frequency or a propensity, $P(E|B) = 1$. If we are to avoid trivialization in many important cases, the expectedness must be treated as a personal probability. To anyone who, like me, wants to base scientific theory preference or choice on objective considerations, this result poses a serious problem.

The net result is a twofold problem. First, by focusing on the expectedness, we do *not* escape the need to deal explicitly with the likelihoods. In [section] 6 I shall discuss the difficulties that arise when we focus on the likelihoods, especially the problem of the likelihood on the catchall hypothesis. Second, the expectedness defies interpretation as an objective probability. In [section] 7 I shall propose a strategy for avoiding involvement with either the expectedness or the likelihood on the catchall. That maneuver will, I hope, keep open the possibility of an objective basis for the evaluation of scientific hypotheses.

6 | Likelihoods

Equations (1), (3), and (4) are different forms of Bayes's theorem, and each of them contains a likelihood, $P(E|T.B)$, in the numerator. Two trivial cases can be noted at the outset. First, if the conjunction of theory T and background knowledge B are logically incompatible with evidence E , the likelihood equals zero, and the posterior probability, $P(T|E.B)$, automatically becomes zero.²⁷ Second, as we have already noticed, if $T.B$ entails E , that likelihood equals one, and consequently drops out, as in equation (5).

Another easy case occurs when the hypothesis T involves various kinds of randomness assumptions, for example, the independence of a series of trials on a chance setup.²⁸ Consider, for example, the case of a coin that has been tossed 100 times, with the result that heads showed in 63 cases and tails in 37. We assume that the tosses are independent, but we are concerned whether the system consisting of the coin and tossing mechanism is biased. Calculation shows that the probability, given an unbiased coin and tossing mechanism, of the actual frequency of heads differing

from $\frac{1}{2}$ by 20 percent or more on 100 tosses (i.e., falling outside of the range 40 to 60) is about .05. Thus, the likelihood of the outcome on the hypothesis that the coin and mechanism are fair is less than .05. On the hypothesis that the coin has a 60 to 40 bias for heads, by contrast, the probability that the number of heads in 100 trials differs from $\frac{60}{100}$ by less than 20 percent (i.e., lies within the 48 to 72 range) is well above .95. These are the kinds of likelihoods that would be used to compare the *null hypothesis* that the coin is fair with the hypothesis that it has a certain bias.²⁹ This example typifies a wide variety of cases, including the above-mentioned controlled experiment on rats and saccharin, in which statistical significance tests are applied. These yield a comparison between the probability of the observed result if the hypothesis is correct and the probability of the same result on a null hypothesis.

In still another kind of situation the likelihood $P(E|T.B)$ is straightforward. Consider, for example, the case in which a physician takes an X-ray for diagnostic purposes. Let T be the hypothesis that the patient has a particular disease and let E be a certain appearance on the film. From long medical experience it may be known that E occurs in 90 percent of all cases in which that disease is present. In many cases, as this example suggests, there may be accumulated frequency data from which the value of $P(E|T.B)$ can be derived.

Unfortunately, life with likelihoods is not always as simple as the foregoing cases suggest. Consider an important case, which I will present in a highly unhistorical way. In comparing the Copernican and Ptolemaic cosmologies, it is easy to see that the phases of Venus are critical. According to the Copernican system, Venus should exhibit a broad set of phases from a narrow crescent to an almost full disk. According to the Ptolemaic system, Venus should always present nearly the same crescent-shaped appearance. One of Galileo's celebrated telescopic observations was of the phases of Venus. The likelihood of such evidence on the Copernican system is unity; on the Ptolemaic it is zero. This is the decisive sort of case that we cherish.

The Copernican system did, however, face one serious obstacle. On the Ptolemaic system, because the earth does not move, the fixed stars should not appear to change their positions. On the Copernican system, because the earth makes an annual trip around the sun, the fixed stars should appear to change their positions in the course of the year. The very best astronomical observations, including those of Tycho Brahe, failed to reveal any observable stellar parallax.³⁰ However, it was realized that, if the fixed stars are at a very great distance from the earth, stellar parallax, though real, would be too small to be observed. Consequently, the likelihood $P(E|T.B)$, where T is the Copernican system and E the absence of observable stellar parallax, is not zero. At the time of the scientific revolution, prior to the advent of Newtonian mechanics, there seemed no reasonable way to evaluate this likelihood. The assumption that the fixed

stars are almost unimaginably distant from the earth was a highly ad hoc, and consequently implausible, auxiliary hypothesis to adopt just to save the Copernican system. Among other things, Christians did not like the idea that heaven was so very far away.

The most reasonable resolution of this anomaly was offered by Tycho Brahe, whose cosmology placed the earth at rest, with the sun and moon moving in orbits around the earth, but with all of the other planets moving in orbits around the sun. In this way both the observed phases of Venus and the absence of observable stellar parallax could be accommodated. Until Newton's dynamics came upon the scene, it seems to me, Tycho's system was clearly the best available theory.

In [section] 2 I suggested that the following form of Bayes's theorem is the most appropriate for use in actual scientific cases in which more than one hypothesis is available for serious consideration:

$$P(T_i|B.E) = \frac{P(T_i|B) \times P(E|T_i.B)}{\sum_{i=1}^k [P(T_i|B) \times P(E|T_i.B)]} \quad 4$$

It certainly fits the foregoing example in which we compared the Ptolemaic, Copernican, and Tychonic systems. This equation involves a mutually exclusive and exhaustive set of hypotheses T_1, \dots, T_{k-1}, T_k , where $T_1 - T_{k-1}$ are seriously entertained and T_k is the catchall. Thus, the scientist who wants to calculate the posterior probability of one particular hypothesis T_i on the basis of evidence E must ascertain likelihoods of three types: (1) the probability of evidence E given T_i , (2) the probability of that evidence on each of the other seriously considered alternatives T_j ($j \neq i$, $j \neq k$), and (3) the probability of that evidence on the catchall T_k .

In considering the foregoing example, I suggested that, although likelihoods in the first two categories are sometimes straightforward, there are cases in which they turn out to be quite problematic. We shall look at more examples in which they present difficulties as our discussion proceeds. But the point to be emphasized right now is the utter intractability of the likelihood on the catchall. The reason for this difficulty is easy to see. Whereas the seriously considered candidates are bona fide hypotheses, the catchall is a hypothesis only in a Pickwickian sense. It refers to all of the hypotheses we are *not* taking seriously, including all those that have not been thought of as yet; indeed, the catchall is logically equivalent to their disjunction. These will often include brilliant discoveries in the future history of science that will eventually solve our most perplexing problems.

Among the hypotheses hidden in the catchall are some that, in conjunction with present available background information, entail the present evidence E . On such as-yet-undiscovered hypotheses the likelihood is one. Obviously, however, the fact that its probability on one particular hypoth-

esis is unity does not entail anything about its probability on some disjunction containing that hypothesis as one of its disjuncts. These considerations suggest to me that the likelihood on the catchall is totally intractable. To try to evaluate the likelihood on the catchall involves, it seems to me, an attempt to guess the future history of science. That is something we cannot do with any reliability.

In any situation in which a small number of theories are competing for ascendancy it is tempting, though quite illegitimate, simply to ignore the likelihood on the catchall. In the nineteenth century, for instance, scientists asked what the probability of a given phenomenon is on the wave theory of light and what it is on the corpuscular theory. They did not seriously consider its probability if neither of these theories is correct. Yet we see, from the various forms in which Bayes's theorem is written, that either the expectedness or the likelihood on the catchall is an indispensable ingredient. In the next section I shall offer a *legitimate* way of eliminating those probabilities from our consideration.

7 | Choosing Between Theories

Kuhn has often maintained that in actual science the problem is never to evaluate one particular hypothesis or theory in isolation; it is always a matter of choosing from among two or more viable alternatives. He has emphasized that an old theory is never completely abandoned unless there is currently available a rival to take its place. Given that circumstance, it is a matter of choosing between the old and the new. On this point I think that Kuhn is quite right, especially as regards reasonably mature sciences. And this insight provides a useful clue on how to use Bayes's theorem to explicate the logic of scientific confirmation.

Suppose that we are trying to choose between T_1 and T_2 , where there may or may not be other serious alternatives in addition to the catchall. By letting $i = 1$ and $i = 2$, we can proceed to write equation (4) for each of these candidates. Noting that the denominators of the two are identical, we can form their ratio as follows:

$$\frac{P(T_1|E.B)}{P(T_2|E.B)} = \frac{P(T_1|B) \times P(E|T_1.B)}{P(T_2|B) \times P(E|T_2.B)} \quad 6$$

No reference to the catchall hypothesis appears in this equation. Since the catchall is not a bona fide hypothesis, it is not a contender, and we need not try to calculate its posterior probability. The use of equation (6) frees us from the need to deal either with the expectedness of E or with its probability on the catchall.

Equation (6) yields a relation that can be regarded as a *Bayesian*

algorithm for theory preference. Suppose that, prior to the emergence of evidence E , you prefer T_1 to T_2 ; that is, $P(T_1|B) > P(T_2|B)$. Then E becomes available. You should change your preference in the light of E if and only if $P(T_2|E.B) > P(T_1|E.B)$. From (6) it follows that

$$P(T_2|E.B) > P(T_1|E.B) \text{ iff } P(E|T_2.B)/P(E|T_1.B) > P(T_1|B)/P(T_2|B). \quad 7$$

In other words, you should change your preference to T_2 if the ratio of the likelihoods is greater than the reciprocal of the ratio of the respective prior probabilities. A corollary is that, if both $T_1.B$ and $T_2.B$ entail E , so that:

$$P(E|T_1.B) = P(E|T_2.B) = 1,$$

the occurrence of E can never change the preference rating between the two competing theories.

At the end of [section] 4 I made reference to the well-known phenomenon of washing out of priors in connection with the use of Bayes's theorem. One might well ask what happens to this swamping when we switch from Bayes's theorem to the ratio embodied in equation (6).³¹ The best answer, I believe, is this. If we are dealing with two hypotheses that are serious contenders in the sense that they do not differ too greatly in plausibility, the ratio of the priors will be of the order of unity. If, as the observational evidence accumulates, the likelihoods come to differ greatly, the ratio of the likelihoods will swamp the ratio of the priors. Recall the example of the tossed coin. Suppose we consider the prior probability of a fair device to be ten times as large as that of a biased device. If about the same proportion of heads occurs in 500 tosses as occurred in the aforementioned 100, the likelihood on the null hypothesis would be virtually zero and the likelihood on the hypothesis that the device has a bias approximating the observed frequency would be essentially indistinguishable from unity. The ratio of prior probabilities would obviously be completely dominated by the likelihood ratio.

8 | Plausible Scenarios

Although, by appealing to equation (6), we have eliminated the need to deal with the expectedness or the likelihood on the catchall, we cannot claim to have dealt adequately with the likelihoods on the hypotheses we are seriously considering, for their values are not always straightforwardly ascertainable. We have already mentioned one example, namely, the probability of absence of observable stellar parallax on the Copernican hypothesis. We noted that, by adding an auxiliary hypothesis to the effect that

the fixed stars are located an enormous distance from the Earth, we could augment the Copernican hypothesis in such a way that the likelihood on this augmented hypothesis is one. But, for many reasons, this auxiliary assumption could hardly be considered plausible in that historical context. By now, of course, we have measured the parallax of relatively nearby stars, and from those values have calculated these distances. They are extremely far from us in comparison to the familiar objects in our solar system.

Consider another well-known example. During the seventeenth and eighteenth centuries the wave and corpuscular theories of light received considerable scientific attention. Each was able to explain certain important optical phenomena, and each faced fundamental difficulties. The corpuscular hypothesis easily explained how light could travel vast distances through empty space, and it readily explained sharp shadows. The theory of light as a longitudinal wave explained various kinds of diffraction phenomena, but failed to deal adequately with polarization. When, early in the nineteenth century, light was conceived as a transverse wave, the wave theory explained polarization as well as diffraction quite straightforwardly. And Huygens had long since shown how the wave theory could handle rectilinear propagation and sharp shadows. For most of the nineteenth century the wave theory dominated optics.

The proponent of the particle theory could still raise a serious objection. What is the likelihood of a wave propagating in empty space? Lacking a medium, the answer is zero. So wave theorists augmented their theory with the auxiliary assumption that all of space is filled with a peculiar substance known as the *luminiferous ether*. This substance was postulated to have precisely the properties required to transmit light waves.

The process I have been describing can appropriately be regarded as the discovery and introduction of *plausible scenarios*. A theory is confronted with an *anomaly*—a phenomenon that appears to have a small, possibly zero, likelihood given that theory. Proponents of the theory search for some auxiliary hypothesis that, if conjoined to the theory, renders the likelihood high, possibly unity. This move shifts the burden of the argument to the plausibility of the new auxiliary hypothesis. I mentioned two instances involved in the wave theory of light. The first was the auxiliary assumption that the wave is transverse. This modification of the theory was sufficiently plausible to be incorporated as an integral part of the theory. The second was the luminiferous ether. The plausibility of this auxiliary hypothesis was debated throughout the nineteenth, and into the twentieth, century. The ether had to be dense enough to transmit transverse waves (which require a denser medium than do longitudinal waves) and thin enough to allow astronomical bodies to move through it without noticeable diminution of speed. Attempts to detect the motion of the earth relative to the ether were unsuccessful. The Lorentz-Fitzgerald contraction hypothesis was an attempt to save the ether theory—that is, another attempt

at a plausible scenario—but it was, of course, abandoned in favor of special relativity.

I am calling these auxiliaries *scenarios* because they are stories about how something could have happened, and *plausible* because they must have some degree of acceptability if they are to be of any help in handling problematic phenomena. The wave theory could handle the Poisson bright spot by deducing it from the theory. There seemed to be no plausible scenario available to the particle theory that could deal with this phenomenon. The same has been said with respect to Foucault's demonstration that the velocity of light is greater in air than it is in water.³²

One nineteenth century optician of considerable importance who did not adopt the wave theory, but remained committed to the Newtonian emission theory, was David Brewster.³³ In a "Report on the Present State of Physical Optics," presented to the British Association for the Advancement of Science in 1831, he maintained that the undulatory theory is "still burthened with difficulties and cannot claim our implicit assent."³⁴ Brewster freely admitted the unparalleled explanatory and predictive success of the wave theory; nevertheless, he considered it false.

Among the difficulties Brewster found with the wave theory, two might be mentioned. First, he considered the wave theory implausible, for the reason that it required "an *ether* invisible, intangible, imponderable, inseparable from all bodies, and extending from our own eye to the remotest verge of the starry heavens."³⁵ History has certainly vindicated him on that issue. Second, he found the wave theory incapable of explaining a phenomenon that he had discovered himself, namely, *selective absorption*—dark lines in the spectrum of sunlight that has passed through certain gases. Brewster points out that a gas may be opaque to light of one particular index of refraction in flint glass, while transmitting freely light whose refractive indices in the same glass are only the tiniest bit higher or lower. Brewster maintained that there was no plausible scenario the wave theorists could devise that would explain why the ether permeating the gas transmits two waves of very nearly the same wavelength, but does not transmit light of a very precise wavelength lying in between.

There is no fact analogous to this in the phenomena of sound, and I can form no conception of a simple elastic medium so modified by the particles of the body which contains it, as to make such an extraordinary selection of the undulations which it stops or transmits. . . .³⁶

Brewster never found a plausible scenario by means of which the Newtonian theory he favored could cope with absorption lines, nor could proponents of the wave theory find one to bolster their viewpoint. Dark absorption lines remained anomalous for both the wave and particle theories; neither could see a way to furnish them with high likelihood.

With hindsight we can say that the catchall hypothesis was looking

very strong at this point. We recognize that the dark absorption lines in the spectrum of sunlight are closely related to the discrete lines in the emission spectra of gases, and that they, in turn, are intimately bound up with the problem of the stability of atoms. These phenomena played a major role in the overthrow of classical physics at the turn of the twentieth century.

I have introduced the notion of a plausible scenario to deal with problematic likelihoods. Likelihoods can cause trouble for a scientific theory for either of two reasons. First, if you have a pet theory that confers an extremely small—for all practical purposes zero—likelihood on some observed phenomenon, that is a problem for that favored theory. You try to come up with a plausible scenario according to which the likelihood will be larger—ideally, unity. Second, if there seems to be no way to evaluate the likelihood of a piece of evidence with respect to some hypothesis of interest, that is another sort of problem. In this case, we search for a plausible scenario that will make the likelihood manageable, whether this involves assigning it a high, medium, or low value.

What does this mean in terms of the Bayesian approach I am advocating? Let us return to:

$$\frac{P(T_1|E.B)}{P(T_2|E.B)} = \frac{P(T_1|B) \times P(E|T_1.B)}{P(T_2|B) \times P(E|T_2.B)}, \quad 6$$

which contains two likelihoods. Suppose, as in nineteenth-century optics, that both likelihoods are problematic. As we have seen, we search for plausible scenarios A_1 and A_2 to augment T_1 and T_2 respectively. If the search has been successful, we can assess the likelihoods of E with respect to the augmented theories $A_1.T_1$ and $A_2.T_2$. Consequently, we can modify (6) so as to yield

$$\frac{P(A_1.T_1|E.B)}{P(A_2.T_2|E.B)} = \frac{P(A_1.T_1|B) \times P(E|A_1.T_1.B)}{P(A_2.T_2|B) \times P(E|A_2.T_2.B)}. \quad 8$$

In order to use this equation to compare the posterior probabilities of the two augmented theories, we must assess the plausibilities of the scenarios, for the prior probabilities of both augmented theories— $A_1.T_1$ and $A_2.T_2$ —appear in it. In [section] 4 I tried to explain how prior probabilities can be handled—that is, how we can obtain at least rough estimates of their values. If, as suggested, the plausible scenarios have made the likelihoods ascertainable, then we can use them in conjunction with our determinations of the prior probabilities to assess the ratio of the posterior probabilities. We have, thereby, handled the central issue raised by Kuhn, namely, what is the basis for preference between two theories.³⁷ Equation (8) is a Bayesian algorithm.

If either augmented theory, in conjunction with background knowledge B , entails E , then the corresponding likelihood is one and it drops out of (8). If both likelihoods drop out we have the special case in which

$$\frac{P(A_1.T_1|E.B)}{P(A_2.T_2|E.B)} = \frac{P(A_1.T_1|B)}{P(A_2.T_2|B)}, \quad 9$$

thereby placing the *whole* burden on the prior probabilities—the plausibility considerations. Equation (9) represents a simplified Bayesian algorithm that is applicable in this type of special case.

Another type of special case was mentioned above. If, as in our coin-tossing example, the values of the prior probabilities do not differ drastically from one another, but the likelihoods become widely divergent as the observational evidence accumulates, there will be a washing out of the priors. In this case, the ratio of the posterior probabilities equals, for practical purposes, the ratio of the likelihoods.

The use of either (8) or (9) as an algorithm for theory choice does not imply that all scientists will agree on the numerical values or prefer the same theory. The evaluation of prior probabilities clearly demands the kind of scientific judgment whose importance Kuhn has rightly insisted upon. It should also be clearly remembered that these formulas provide no evaluations of individual theories; they furnish only comparative evaluations. Thus, instead of yielding a prediction regarding the chances of one particular theory being a component of “completed science,” they compare existing theories with regard to their present merits.

9 | Kuhn's Criteria

Early in this paper I quoted five criteria that Kuhn mentioned in connection with his views on the rationality and objectivity of science. The time has come to relate them explicitly to the Bayesian approach I have been attempting to elaborate. In order to appreciate the significance of these criteria it is important to distinguish three aspects of scientific theories that may be called *informational virtues*, *confirmational virtues*, and *economic virtues*. Up to this point we have concerned ourselves almost exclusively with confirmation, for our use of Bayes's theorem is germane only to the confirmational virtues. But since Kuhn's criteria patently refer to the other virtues as well, we must also say a little about them.

Consider, for example, the matter of *scope*. Newton's three laws of motion and his law of universal gravitation obviously have greater scope than the conjunction of Galileo's law of falling bodies and Kepler's three laws of planetary motion. This means, simply, that Newtonian mechanics contains more information than the laws of Kepler and Galileo taken

together. Given a situation of this sort, we prefer the more informative theory because it is a basic goal of science to increase our knowledge as much as possible. We might, of course, hesitate to choose a highly informative theory if the evidence for it were extremely limited or shaky, because the desire to be right might overrule the desire to have more information content. But in the case at hand that consideration does not arise.

In spite of its intuitive attraction, however, the appeal to scope is not altogether unproblematic. There are two ways in which we might construe the Galileo-Kepler-Newton example of the preceding paragraph. First, we might ignore the small corrections mandated by Newton's theory in the laws of Galileo and Kepler. In that case we can clearly claim greater scope for Newton's laws than for the conjunction of Galileo's and Kepler's laws, since the latter is entailed by the former but not conversely. Where an entailment relation holds we can make good sense of comparative scope.

Kuhn, however, along with most of the historically oriented philosophers, has been at pains to deny that science progresses by finding more general theories that include earlier theories as special cases. Theory choice or preference involves *competing* theories that are *mutually incompatible* or *mutually incommensurable*. To the best of my knowledge Kuhn has not offered any precise characterization of scope; Karl Popper, in contrast, has made serious attempts to do so. In response to Popper's efforts, Adolf Grünbaum has effectively argued that none of the Popperian measures can be usefully applied to make comparisons of scope among mutually incompatible competing theories.³⁸ Consequently, the concept of scope requires fundamental clarification if we are to use it to understand preferences among competing theories. However, since scope refers to information rather than confirmation, it plays no role in the Bayesian program I have been endeavoring to explicate. We can thus put aside the problem of explicating that difficult concept.

Another of Kuhn's criteria is *accuracy*. It can, I think, be construed in two different ways. The first has to do with informational virtues; the second with economic. On the one hand, two theories might both make true predictions regarding the same phenomena, but one of them might give us precise predictions where the other gives only predictions that are less exact. If, for example, one theory enables us to predict that there will be a solar eclipse on a given day, and that its path of totality will cross North America, it may well be furnishing correct information about the eclipse. If another theory gives not only the day, but also the time, and not only the continent, but also the precise boundaries, the second provides much more information, at least with respect to this particular occurrence. It is not that either is incorrect; rather, the second yields more knowledge than the first. However, it should be clearly noted—as it was in the case of scope—that these theories are not incompatible or incommensurable competitors (at least with respect to this eclipse), and hence

do not illustrate the interesting type of theory preference with which Kuhn is primarily concerned.

On the other hand, one theory may yield predictions that are nearly, but not quite, correct, while another theory yields predictions that are entirely correct—or, at least, more nearly correct. Newtonian astrophysics does well in ascertaining the orbit of the earth, but general relativity introduces a correction of 3.8 seconds of arc per century in the precession of its perihelion.³⁹ Although the Newtonian theory is literally false, it is used in contexts of this sort because its inaccuracy is small, and the economic gain involved in using it instead of general relativity (the saving in computational effort) is enormous.

The remaining three criteria are *simplicity*, *consistency*, and *fruitfulness*; all of them have direct bearing upon the confirmational virtues. In the treatment of prior probabilities in [section] 4, I briefly mentioned simplicity as a factor having a significant bearing upon the plausibility of theories. More examples could be added, but I think the point is clear.

In the same section I also made passing reference to consistency, but more can profitably be said on that topic. Consistency has two aspects, internal consistency of a theory and its compatibility with other accepted theories. While scientists may be fully justified in *entertaining* collections of statements that contain contradictions, the goal of science is surely to accept only logically consistent theories.⁴⁰ The discovery of an internal inconsistency has a distinctly adverse effect on the prior probability of that theory, to wit, it must go straight to zero.

When we consider the relationships of a given theory to other accepted theories we again find two aspects. There are *deductive* relations of entailment and incompatibility, and there are *inductive* relations of fittingness and incongruity. The deductive relations are quite straightforward. Incompatibility with an accepted theory makes for implausibility; being a logical consequence of an accepted theory makes for a high prior probability. Although deductive subsumption of narrower theories under broader theories is probably something of an oversimplification of actual cases, nevertheless, the ability of an overarching theory to deductively unify diverse domains furnishes a strong plausibility argument.

When it comes to the inductive relations among theories, analogy is, I think, the chief consideration. I have already mentioned the use of analogy in inductively transferring results of experiments from rats to humans. In archaeology, the method of ethnographic analogy, which exploits similarities between extant primitive societies and prehistoric societies, is widely used. In physics, the analogy between the inverse square law of electrostatics and the inverse square law of gravitation provides an example of an important plausibility consideration.

Kuhn's criteria of consistency (broadly construed) and simplicity seem clearly to pertain to assessments of the prior probabilities of theories. They cry out for a Bayesian interpretation.

The final criterion in Kuhn's list is *fruitfulness*; it has many aspects. Some theories prove fruitful by unifying a great many apparently different phenomena in terms of a few simple principles. The Newtonian synthesis is, perhaps, the outstanding example, Maxwellian electrodynamics is also an excellent case. As I suggested above, this ability to accommodate a wide variety of facts tends to enhance the prior probability of a given theory. To attribute diverse success to happenstance, rather than basic correctness, is implausible.

Another sort of fertility involves the predictability of theretofore unknown phenomena. We might mention as familiar illustrations the prediction of the Poisson bright spot by the wave theory of light and the prediction of time dilation by special relativity. These are the kinds of instances in which, in an important sense, the expectedness is low. As we have noted, a small expectedness tends to increase the posterior probability of a hypothesis.

A further type of fertility relates directly to plausible scenarios; a theory is fruitful in this way if it successfully copes with difficulties with the aid of suitable auxiliary assumptions. Newtonian mechanics again provides an excellent example. The perturbations of Uranus were explained by postulating Neptune. The perturbations of Neptune were explained by postulating Pluto.⁴¹ The motions of stars within galaxies and of galaxies within clusters are explained in terms of *dark matter*, concerning which there are many current theories. A theory that readily gives rise to plausible scenarios to deal with problematic likelihoods can boast this sort of fertility.

The discussion of Kuhn's criteria in this section is intended to show how adequately they can be understood within a Bayesian framework—insofar as they are germane to confirmation. If it is sound, we have constructed a fairly substantial bridge connecting Kuhn's views on theory choice with those of the logical empiricists—at least, those who find in Bayes's theorem a suitable schema for characterizing the confirmation of hypotheses and theories.

10 | Rationality vs. Objectivity

In the title of this essay I have used both the concept of *rationality* and that of *objectivity*. It is time to say something about their relationship. Perhaps the best way to approach the distinction between them is to enumerate various grades of rationality. In a certain sense one can be rational without paying any heed at all to objectivity. It is essentially a matter of good housekeeping as far as one's beliefs and degrees of confidence are concerned. As Bayesians have often emphasized, it is important to avoid logical contradictions in one's beliefs and to avoid probabilistic incoherence in one's degrees of conviction. If contradiction or incoherence are

discovered, they must somehow be eliminated; the presence of either constitutes a form of irrationality. But the removal of such elements of irrationality can be accomplished without any appeal to facts outside of the subject's corpus of beliefs and degrees of confidence. To achieve this sort of rationality is to achieve a minimal standard that I have elsewhere called *static rationality*.⁴²

One way in which additional facts may enter the picture is via Bayes's theorem. We have a theory *T* in which we have a particular degree of confidence. A new piece of evidence turns up—some objective fact *E* of which we were previously unaware—and we use Bayes's theorem to calculate a posterior probability of *T*. To accept this value of the posterior probability as one's degree of confidence in *T* is known as *Bayesian conditionalization*. Use of Bayes's theorem does not, however, guarantee objectivity. If the resulting posterior probability of *T* is one we are not willing to accept, we can make adjustments elsewhere to avoid incoherence. After all, the prior probabilities and likelihoods are simply personal probabilities, so they can be adjusted to achieve the desired result. If, however, the requirement of *Bayesian conditionalization* is added to those of static rationality we have a stronger type of rationality that I have called *kine-matic*.⁴³

The highest grade of rationality—what I have called *dynamic rationality*—requires much fuller reference to objective fact than is demanded by advocates of personalism. The most obvious way to inject a substantial degree of objectivity into our deliberations regarding choices of scientific theories is to provide an objective interpretation of the probabilities in Bayes's theorem. Throughout this discussion I have adopted that approach as thoroughly as possible. For instance, I have argued that prior probabilities can be given an objective interpretation in terms of frequencies of success. I have tried to show how likelihoods could be objective—by virtue of entailment relations, tests of statistical significance, or observed frequencies. When the likelihoods created major difficulties, I appealed to plausible scenarios. The result was that an intractable likelihood could be exchanged for a tractable prior probability—namely, the prior probability of a theory in conjunction with an auxiliary assumption.

We noted that the denominators of the right-hand sides of the various versions of Bayes's theorem—equations (1), (3), and (4)—contain either an expectedness or a likelihood on the catchall. It seems to me futile to try to construe either of these probabilities objectively. Consequently, in [section] 7 I introduced equation (6), which involves a ratio of two instances of Bayes's theorem, and from which the expectedness and the likelihood on the catchall drop out. Confining our attention, as Kuhn recommends, to comparing the merits of competing theories, rather than offering absolute evaluations of individual theories, we were able to eliminate the probabilities that most seriously defy objective interpretation.

11 | Conclusions

For many years I have been convinced that plausibility arguments in science have constituted a major stumbling block to an understanding of the logic of scientific inference. Kuhn was not alone, I believe, in recognizing that considerations of plausibility constitute an essential aspect of scientific reasoning, without seeing where they fit into the logic of science. If one sees confirmation solely in terms of the crude hypothetico-deductive method, there is no place for them. There is, consequently, an obvious incentive for relegating plausibility considerations to heuristics. If one accepts the traditional distinction between the *context of discovery* and the *context of justification*, it is tempting to place them in the former context. But Kuhn recognized, I think, that plausibility arguments enter into the justifications of choices of theories, with the result that he became skeptical of the value of that distinction. If, as I believe, plausibility considerations are simply evaluations of prior probabilities of hypotheses or theories, then it becomes apparent via Bayes's theorem that they play an indispensable role in the context of justification. We do not need to give up that important distinction.

At several places in this paper I have spoken of Bayesian algorithms, mainly because Kuhn introduced that notion into the discussion. I have claimed that such algorithms exist—and attempted to exhibit them—but I accord *very little* significance to that claim. The algorithms are trivial, what is important is the scientific judgment involved in assessing the probabilities that are fed into the equations. The algorithms give frameworks in terms of which to understand the role of the sort of judgment upon which Kuhn rightly placed great emphasis.

The history of science chronicles the successes and failures of attempts at scientific theorizing. If the Bayesian analysis I have been offering is at all sound, history of science—in addition to contemporary scientific experience, of course—provides a rich source of information relevant to the prior probabilities of the theories among which we are at present concerned to make objective and rational choices. This viewpoint captures, I believe, the point Kuhn made at the beginning of his first book:

But an age as dominated by science as our own does need a perspective from which to examine the scientific beliefs which it takes so much for granted, and history provides one important source of such perspective. If we can discover the origins of some modern scientific concepts and the way in which they supplanted the concepts of an earlier age, we are more likely to evaluate intelligently their chances for survival.⁴⁴

I suggested at the outset that an appeal to Bayesian principles could provide some aid in bridging the gap between Hempel's logical-empiricist

approach and Kuhn's historical approach. I hope I have offered a convincing case. However that may be, there remain many unresolved issues. For instance, I have not even broached the problem of incommensurability of paradigms or theories. This is a major issue. For another example, I have assumed uncritically throughout the discussion that the various parties to disputes about theories share a common body *B* of background knowledge. It is by no means obvious that this is a tenable assumption. No doubt other points for controversy remain. I do not for a moment maintain that complete consensus would be in the offing even if both camps were to buy the Bayesian line I have been peddling. But I do hope that some areas of misunderstanding have been clarified.⁴⁵

■ | Notes

1 Thomas S. Kuhn, *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1962; 2d ed., 1970). "Postscript—1969," added to the second edition, contains discussions of some of the major topics that are treated in the present essay.

2 Such philosophers are often characterized by their opponents as *logical positivists*, but this is an egregious historical inaccuracy. Although some of them had been members of or closely associated with the Vienna Circle in their earlier years, none of them retained the early positivistic commitment to phenomenalism and/or instrumentalism in their more mature writings. Reichenbach and Feigl, for example, were outspoken realists, and Carnap regarded physicalism as a tenable philosophical framework. Reichenbach never associated himself with positivism; indeed, he regarded his 1938 book, *Experience and Prediction* (Chicago: University of Chicago Press), as a refutation of logical positivism. I could go on and on.

3 "Symposium: The Philosophy of Carl G. Hempel," *Journal of Philosophy* LXXX, 10 (Oct., 1983), 555–72.

4 I had offered a similar suggestion in "Bayes's Theorem and the History of Science," in *Minnesota Studies in the Philosophy of Science*, vol. 5 *Historical and Philosophical Perspectives of Science*, ed. Roger H. Stuewer (Minneapolis: University of Minnesota Press, 1970), 68–86.

5 Thomas S. Kuhn, *The Essential Tension* (Chicago: University of Chicago Press, 1977), 320–39. The response is given in greater detail in ["Objectivity, Value Judgment, and Theory Choice" (reprinted in chapter 2, this volume)] than it is in the Postscript.

6 "Objectivity, Value Judgment, and Theory Choice," 321–22 [p. 103 above, this volume].

7 Ibid., p. 322 [pp. 103–4 above, this volume].

8 Throughout this paper I shall use the terms "hypothesis" and "theory" more or less interchangeably. Kuhn tends to prefer "theory," while I tend to prefer "hypothesis," but nothing of importance hinges on this usage here.

- 9 As Adolf Grünbaum pointed out to me, if we assume that in a given day the actual frequency of defective can openers produced by the two machines matches precisely the respective probabilities, we can calculate the result as follows. The new machine produces 50 defective can openers and the old machine produces 30, so that 50 out of a total of 80 are produced by the new machine. However, it would be *incorrect* to assume that the frequencies match the probabilities each day; in fact, the probability of an exact match is quite small
- 10 "Objectivity, Value Judgment, and Theory Choice," 328 [p. 109 above, this volume].
11. I remarked above that three probabilities are required to calculate the posterior probability—a prior probability and two likelihoods. Obviously, in view of (2), the theorem on total probability, if we have a prior probability, one of the likelihoods, and the expectedness, we can compute the other likelihood, likewise, if we have one prior probability and both likelihoods, we can compute the expectedness
- 12 *The Foundations of Scientific Inference* (Pittsburgh: University of Pittsburgh Press, 1967), chap. 7
13. A set of degrees of conviction is coherent provided that its members do not violate any of the conditions embodied in the mathematical calculus of probability
- 14 From "Autobiographical Notes," in Paul A. Schilpp, ed., *Albert Einstein Philosopher-Scientist* (Evanston, Ill.: Library of Living Philosophers, 1949), 21–22.
15. This example is discussed in Ronald N. Giere, *Understanding Scientific Reasoning*, 2d ed. (New York: Holt, Rinehart, and Winston, 1984), 274–76
16. I reject the so-called propensity interpretation of probability because, as Paul Humphreys pointed out, the probability calculus accommodates inverse probabilities of the type that occur in Bayes's theorem, but the corresponding inverse propensities do not exist. In the example of the can opener factory, each machine has a certain propensity to produce defective can openers, but it does not make sense to speak of the propensity of a given defective can opener to have been produced by the new machine.
17. A so-called Dutch book is a combination of bets such that, no matter what the outcome of the event upon which the wagers are made, the subject is bound to suffer a net loss.
18. "Truth and Probability," in Frank Plumpton Ramsey, *The Foundations of Mathematics*, ed. R. B. Braithwaite (New York: Humanities Press, 1950).
19. Gerald Holton and Stephen G. Brush, *Introduction to Concepts and Theories in Physical Science*, 2d ed. (Reading, Mass.: Addison-Wesley, 1973), 416, italics in original
20. Abner Shimony, "Scientific Inference," in Robert G. Colodny, ed., *The Nature and Function of Scientific Theories* (Pittsburgh: University of Pittsburgh Press, 1970), 79–172.
21. Martin Gardner, *Fads and Fallacies in the Name of Science* (New York: Dover Publications, 1957), 7–15.
22. Doubleday & Company, 1950.

23. Haim Harari, "The Structure of Quarks and Leptons," *Scientific American* 248, (April 1983): 56–68.
24. *Scientific American*, 253 (Sept. 1985): 116. Adolf Grünbaum, *The Foundations of Psychoanalysis* (Berkeley/Los Angeles/London: University of California Press, 1984), 202–4, criticizes Motley's account of Freud's theory, he considers Motley's version a distortion, and points out that Freud's motivational explanations were explicitly confined to a very circumscribed set of slips. He defends Freud against Motley's criticism on the grounds that Freud's actual account has greater complexity than Motley gives it credit for.
25. "A Bayesian Approach to the Paradoxes of Confirmation," in Jaakko Hintikka and Patrick Suppes, eds., *Aspects of Inductive Logic* (Amsterdam: North Holland, 1966), 202–3.
26. I have discussed the relations between personal probabilities and objective probabilities in "Dynamic Rationality: Propensity, Probability, and Credence," in James H. Fetzer, ed., *Probability and Causality* (Dordrecht: Reidel 1988), 3–40.
27. As Duhem has made abundantly clear, in such cases we may be led to reexamine our background knowledge B, which normally involves auxiliary hypotheses, to see whether it remains acceptable in the light of the negative outcome E. Consequently, refutation of T is not usually as automatic as it appears in the simplified account just given. Nevertheless, the probability relation just stated is correct.
28. Exchangeability is the personalist's surrogate for randomness, it means that the subject would draw the same conclusion regardless of the order in which the members of an observed sample occurred.
29. Note that, in order to get the posterior probability—the probability that the observed results were produced by a biased device—the prior probabilities have to be taken into account.
30. Indeed, stellar parallax was not detected until the nineteenth century.
31. This question was, in fact, raised by Adolf Grünbaum in a private communication.
32. See, for example, Gerald Holton and Stephen G. Brush, *Introduction to Concepts and Theories in Physical Science*, 2d ed. (Reading, Mass.: Addison-Wesley, 1973), 392–93.
33. An excellent account of Brewster's position can be found in John Worrall, "Scientific Revolutions and Scientific Rationality: The Case of the Elderly Hold-out," [in C. Wade Savage, ed., *Scientific Theories*, vol. 14, *Minnesota Studies in the Philosophy of Science* (Minneapolis: University of Minnesota Press, 1990)], pp. 319–36.
34. Quoted by Worrall, p. 321.
35. Quoted by Worrall, p. 322.
36. Quoted by Worrall, p. 323.
37. If more than two theories are serious candidates, the pairwise comparison can be repeated as many times as necessary.

38. Adolf Grünbaum, "Can a Theory Answer More Questions Than One of Its Rivals?" *British Journal for the Philosophy of Science* 27 (1976): 1–23
39. Steven Weinberg, *Gravitation and Cosmology* (New York: John Wiley & Sons, 1972), 198. Note that this correction is smaller by an order of magnitude than the correction of 43 seconds of arc per century for Mercury.
40. See Joel Smith, *The Status of Inconsistent Statements in Scientific Inquiry* (doctoral dissertation, University of Pittsburgh, 1987).
41. Unfortunately, recent evidence regarding the mass of Pluto strongly suggests that Pluto is not sufficiently massive to explain the perturbations of Neptune. A different plausible scenario is needed, but I do not know of any serious candidates that have been offered.
42. See "Dynamic Rationality" (note 26 above), 5–12, for a more detailed discussion of various grades of rationality. The term "static" was chosen to indicate the lack of any principled method for changing personal probabilities in the face of inconsistency or incoherence.
43. *Ibid.*, esp. pp. 11–12.
44. Thomas S. Kuhn, *The Copernican Revolution* (Cambridge: Harvard University Press, 1957), 3–4.
45. I should like to express my deepest gratitude to Adolf Grünbaum and Philip Kitcher for important criticism and valuable suggestions with respect to an earlier version of this paper.

Why I Am Not a Bayesian

The aim of confirmation theory is to provide a true account of the principles that guide scientific argument insofar as that argument is not, and does not purport to be, of a deductive kind. A confirmation theory should serve as a critical and explanatory instrument quite as much as do theories of deductive inference. Any successful confirmation theory should, for example, reveal the structure and fallacies, if any, in Newton's argument for universal gravitation, in nineteenth-century arguments for and against the atomic theory, in Freud's arguments for psychoanalytic generalizations. Where scientific judgments are widely shared, and sociological factors cannot explain their ubiquity, and analysis through the lens provided by confirmation theory reveals no good explicit arguments for the judgments, confirmation theory ought at least sometimes to suggest some good arguments that may have been lurking misperceived. Theories of deductive inference do that much for scientific reasoning insofar as that reasoning is supposed to be demonstrative. We can apply quantification theory to assess the validity of scientific arguments, and although we must almost always treat such arguments as enthymematic, the premises we interpolate are not arbitrary; in many cases, as when the same subject matter is under discussion, there is a common set of suppressed premises. Again, there may be differences about the correct logical form of scientific claims; differences of this kind result in (or from) different formalizations, for example, of classical mechanics. But such differences often make no difference for the assessment of validity in actual arguments. Confirmation theory should do as well in its own domain. If it fails, then it may still be of interest for many purposes, but not for the purpose of understanding scientific reasoning.

The aim of confirmation theory ought not to be simply to provide

precise replacements for informal methodological notions, that is, explanations of them. It ought to do more, in particular, confirmation theory ought to *explain* both methodological truisms and particular judgments that have occurred within the history of science. By "explain" I mean at least that confirmation theory ought to provide a rationale for methodological truisms and ought to reveal some systematic connections among them and, further, ought without arbitrary or question-begging assumptions to reveal particular historical judgments as in conformity with its principles.

Almost everyone interested in confirmation theory today believes that confirmation relations ought to be analyzed in terms of *probability* relations. Confirmation theory is the theory of probability plus introductions and appendices. Moreover, almost everyone believes that confirmation proceeds through the formation of conditional probabilities of hypotheses on evidence. The basic tasks facing confirmation theory are thus just those of explicating and showing how to determine the probabilities that confirmation involves, developing explications of such metascientific notions as "confirmation," "explanatory power," "simplicity," and so on in terms of functions of probabilities and conditional probabilities, and showing that the canons and patterns of scientific inference result. It was not always so. Probabilistic accounts of confirmation really became dominant only after the publication of Carnap's *Logical Foundations of Probability*,² although of course many probabilistic accounts had preceded Carnap's. An eminent contemporary philosopher³ has compared Carnap's achievement in inductive logic with Frege's in deductive logic: just as before Frege there was only a small and theoretically uninteresting collection of principles of deductive inference, but after him the foundation of a systematic and profound theory of demonstrative reasoning, so with Carnap and inductive reasoning. After Carnap's *Logical Foundations*, debates over confirmation theory seem to have focused chiefly on the interpretation of probability and on the appropriate probabilistic explications of various meta-scientific notions. The meta-scientific notions remain controversial, as does the interpretation of probability, although increasingly logical interpretations of probability are giving way to the doctrine that probability is degree of belief.⁴ In very recent years a few philosophers have attempted to apply probabilistic analyses to derive and to explain particular methodological practices and precepts, and even to elucidate some historical cases.

I believe these efforts, ingenious and admirable as many of them are, are nonetheless misguided. For one thing, probabilistic analyses remain at too great a distance from the history of scientific practice to be really informative about that practice, and in part they do so exactly because they are probabilistic. Although considerations of probability have played an important part in the history of science, until very recently explicit probabilistic arguments for the confirmation of various theories, or prob-

abilistic analyses of data, have been great rarities in the history of science. In the physical sciences at any rate, probabilistic arguments have rarely occurred. Copernicus, Newton, Kepler, none of them give probabilistic arguments for their theories, nor does Maxwell or Kelvin or Lavoisier or Dalton or Einstein or Schrödinger or. . . . There are exceptions. Jon Dorling has discussed a seventeenth-century Ptolemaic astronomer who apparently made an extended comparison of Ptolemaic and Copernican theories in probabilistic terms; Laplace, of course, gave Bayesian arguments for astronomical theories. And there are people, Maxwell for example, who scarcely give a probabilistic argument when making a case for or against scientific hypotheses but who discuss *methodology* in probabilistic terms. This is not to deny that there are many areas of contemporary physical science where probability figures large in confirmation, regression analysis is not uncommon in discussions of the origins of cosmic rays, correlation and analysis of variance in experimental searches for gravitational waves, and so on. It is to say that, explicitly, probability is a distinctly minor note in the history of scientific argument.

The rarity of probability considerations in the history of science is more an embarrassment for some accounts of probability than for others. Logical theories, whether Carnap's or those developed by Hintikka and his students, seem to lie at a great distance from the history of science. Still, some of the people working in this tradition have made interesting steps toward accounting for methodological truisms. My own inclination is to believe that the interest such investigations have stems more from the insights they obtain into syntactic versions of structural connections among evidence and hypotheses than to the probability measures they mesh with these insights. Frequency interpretations suppose that for each hypothesis to be assessed there is an appropriate reference class of hypotheses to which to assign it, and the prior probability of the hypothesis is the frequency of true hypotheses in this reference class. The same is true for statements of evidence, whether they be singular or general. The matter of how such reference classes are to be determined, and determined so that the frequencies involved do not come out to be zero, is a question that has only been touched upon by frequentist writers. More to the point, for many of the suggested features that might determine reference classes, we have no statistics and cannot plausibly imagine those who figure in the history of our sciences to have had them. So conceived, the history of scientific argument must turn out to be largely a history of fanciful guesses. Further, some of the properties that seem natural candidates for determining reference classes for hypotheses—simplicity, for example, seem likely to give perverse results. We prefer hypotheses that posit simple relations among observed quantities, and so on a frequentist view should give them high prior probabilities. Yet simple hypotheses, although often very useful approximations, have most often turned out to be literally false.

At present, perhaps the most philosophically influential view of prob-

ability understands it to be degree of belief. The subjectivist Bayesian (hereafter, for brevity, simply Bayesian) view of probability has a growing number of advocates who understand it to provide a general framework for understanding scientific reasoning. They are singularly unembarrassed by the rarity of explicit probabilistic arguments in the history of science, for scientific reasoning need not be explicitly probabilistic in order to be probabilistic in the Bayesian sense. Indeed, a number of Bayesians have discussed historical cases within their framework. Because of its influence and its apparent applicability, in what follows it is to the subjective Bayesian account that I shall give my full attention.

My thesis is several-fold. First, there are a number of attempts to demonstrate a priori the rationality of the restrictions on belief and inference that Bayesians advocate. These arguments are altogether admirable but ought, I shall maintain, to be unconvincing. My thesis in this instance is not a new one, and I think many Bayesians do regard these a priori arguments as insufficient. Second, there are a variety of methodological notions that an account of confirmation ought to explicate, and methodological truisms involving these notions that a confirmation theory ought to explain: for example, variety of evidence and why we desire it, ad hoc hypotheses and why we eschew them, what separates a hypothesis integral to a theory from one "tacked on" to the theory, simplicity and why it is so often admired, why "deoccamized" theories* are so often disdained, what determines when a piece of evidence is relevant to a hypothesis, what, if anything, makes the confirmation of one bit of theory by one bit of evidence stronger than the confirmation of another bit of theory (or possibly the same bit) by another (or possibly the same) bit of evidence. Although there are plausible Bayesian explications of some of these notions, there are not plausible Bayesian explications of others. Bayesian accounts of methodological truisms and of particular historical cases are of one of two kinds: either they depend on general principles restricting prior probabilities or they don't. My claim is that many of the principles proposed by the first kind of Bayesian are either implausible or incoherent, and that, for want of such principles, the explanations the second kind of Bayesians provide for particular historical cases and for truisms of method are chimeras. Finally, I claim that there are elementary but perfectly common features of the relation of theory and evidence that the Bayesian scheme cannot capture at all without serious—and perhaps not very plausible—revision.

It is not that I think the Bayesian scheme or related probabilistic accounts capture nothing. On the contrary, they are clearly pertinent where the reasoning involved is explicitly statistical. Further, the accounts

* Glymour explains the notion of deoccamization later in his article. It means adding to a theory a variable for which there is no independent evidence and whose values can be determined only by combining the values of other quantities.

developed by Carnap, his predecessors, and his successors are impressive systematizations and generalizations, in a probabilistic framework, of certain principles of ordinary reasoning. But so far as understanding scientific reasoning goes, I think it is very wrong to consider our situation to be analogous to that of post-Fregean logicians, our subject matter transformed from a hodgepodge of principles by a powerful theory whose outlines are clear. We flatter ourselves that we possess even the hodgepodge. My opinions are outlandish, I know; few of the arguments I shall present in their favor are new, and perhaps none of them is decisive. Even so, they seem sufficient to warrant taking seriously entirely different approaches to the analysis of scientific reasoning.

The theories I shall consider share the following framework, more or less: There is a class of sentences that express all hypotheses and all actual or possible evidence of interest; the class is closed under Boolean operations. For each ideally rational agent, there is a function defined on all sentences such that, under the relation of logical equivalence, the function is a probability measure on the collection of equivalence classes. The probability of any proposition represents the agent's degree of belief in that proposition. As new evidence accumulates, the probability of a proposition changes according to Bayes' rule: the posterior probability of a hypothesis on the new evidence is equal to the prior conditional probability of the hypothesis on the evidence. This is a scheme shared by diverse accounts of confirmation. I call such theories "Bayesian," or sometimes, "personalist."

We certainly have *grades* of belief. Some claims I more or less believe, some I find plausible and tend to believe, others I am agnostic about, some I find implausible and far-fetched, still others I regard as positively absurd. I think everyone admits some such gradations, although descriptions of them might be finer or cruder. The personalist school of probability theorists claim that we also have *degrees* of belief, degrees that can have any value between 0 and 1 and that ought, if we are rational, to be representable by a probability function. Presumably, the degrees of belief are to covary with everyday gradations of belief, so that one regards a proposition as preposterous and absurd just if his degree of belief in it is somewhere near zero, and he is agnostic just if his degree of belief is somewhere near a half, and so on. According to personalists, then, an ideally rational agent always has his degrees of belief distributed so as to satisfy the axioms of probability, and when he comes to accept a new belief he also forms new *degrees* of belief by conditionalizing on the newly accepted belief. There are any number of refinements, of course, but that is the basic view.

Why should we think that we really do have *degrees* of belief? Personalists have an ingenious answer: people have them because we can measure the degrees of belief that people have. Assume that no one (rational) will accept a wager on which he expects a loss, but anyone (ra-

tional) will accept any wager on which he expects a gain. Then we can measure a person's degree of belief in proposition P by finding, for fixed amount v , the highest amount u such that the person will pay u in order to receive $u + v$ if P is true but receive nothing if P is not true. If u is the greatest amount the agent is willing to pay for the wager, his expected gain on paying u must be zero. The agent's gain if P is the case is v ; his gain if P is not the case is $-u$. Thus*

$$(v) \times \text{Prob}(P) + (-u) \times \text{Prob}(\sim P) = 0$$

Since $\text{Prob}(\sim P) = 1 - \text{Prob}(P)$ we have

$$\text{Prob}(P) = u/(u + v)$$

The reasoning is clear: any sensible person will act so as to maximize his expected gain, thus, presented with a decision whether or not to purchase a bet, he will make the purchase just if his expected gain is greater than zero. So the betting odds he will accept determine his degree of belief.⁵

I think that this device really does provide evidence that we have, or can produce, degrees of belief, in at least some propositions, but at the same time it is evident that betting odds are not an unobjectionable device for the measurement of degrees of belief. Betting odds could fail to measure degrees of belief for a variety of reasons: the subject may not believe that the bet will be paid off if he wins, or he may doubt that it is clear what constitutes winning, even though it is clear what constitutes losing. Things he values other than monetary gain (or whatever) may enter into his determination of the expected utility of purchasing the bet: for example, he may place either a positive or a negative value on risk itself. And the very fact that he is offered a wager on P may somehow change his degree of belief in P .

Let us suppose, then, that we do have degrees of belief in at least some propositions, and that in some cases they can be at least approximately measured on an interval from 0 to 1. There are two questions: why should we think that, for rationality, one's degrees of belief must satisfy the axioms of probability, and why should we think that, again for rationality, changes in degrees of belief ought to proceed by conditionalization? One question at a time. In using betting quotients to measure degrees of belief it was assumed that the subject would act so as to maximize *expected* gain. The betting quotient determined the degree of belief by determining the coefficient by which the gain is multiplied in case that P is true in the expression for the expected gain. So the betting quotient determines a degree of belief, as it were, in the *role* of a probability. But why should

* Here and elsewhere in Glymour's article, we have changed the notation slightly in the interests of stylistic consistency and ease of comprehension.

the things, degrees of belief, that play this role, be probabilities? Suppose that we do choose those actions that maximize the sum of the product of our degrees of belief in each possible outcome of the action and the gain (or loss) to us of that outcome. Why must the degrees of belief that enter into this sum be probabilities? Again there is an ingenious argument. If one acts so as to maximize his expected gain using a degree-of-belief function that is not a probability function, and if for every proposition there were a possible wager (which, if it is offered, one believes will be paid off if it is accepted and won), then there is a circumstance, a combination of wagers, that one would enter into if they were offered, and in which one would suffer a net loss whatever the outcome. That is what the Dutch Book argument shows; what it counsels is prudence.

Some of the reasons why it is not clear that betting quotients are accurate measures of degrees of belief are also reasons why the Dutch Book argument is not conclusive: there are many cases of propositions in which we may have degrees of belief, but on which, we may be sure, no acceptable wager will be offered us; again, we may have values other than the value we place on the stakes and these other values may enter into our determination whether or not to gamble; and we may not have adopted the policy of acting so as to maximize our expected gain or our expected utility: that is, we may save ourselves from having book made against us by refusing to make certain wagers, or combinations of wagers, even though we judge the odds to be in our favor.

The Dutch Book argument does not succeed in showing that in order to avoid absurd commitments, or even the possibility of such commitments, one must have degrees of belief that are probabilities. But it does provide a kind of justification for the personalist viewpoint, for it shows that if one's degrees of belief are probabilities, then a certain kind of absurdity is avoided. There are other ways of avoiding that kind of absurdity, but at least the personalist way is one such.⁶

One of the common objections to Bayesian theory is that it fails to provide any connection between what is inferred and what is the case. The Bayesian reply is that the method guarantees that, in the long run, everyone will agree on the truth. Suppose that B_i are a set of mutually exclusive, jointly exhaustive hypotheses, each with probability $P(B_i)$.^{*} Let \bar{x}_r be a sequence of random variables with a finite set of values and conditional distribution given by $P(\bar{x}_r = x_r | B_i) = \epsilon(x_r | B_i)$; then we can think of the values x_r as the outcomes of experiments, each hypothesis determining a likelihood for each outcome. Suppose that no two hypotheses have the same likelihood distribution, that is, for $i \neq j$ it is not the case that for all values x_r of \bar{x}_r , $\epsilon(x_r | B_i) = \epsilon(x_r | B_j)$, where the ϵ 's are defined as above. Let \bar{x} denote the first n of these variables, where x is a value of \bar{x} .

* In his book, Glymour follows Savage in writing $P(B_i)$ as $B(i)$.

Now imagine an observation of these n random variables. In Savage's words:

Before the observation, the probability that the probability given x of whichever element of the partition actually obtains will be greater than α is

$$\sum_i P(B_i) \times P[(P(B_i|x) > \alpha) | B_i]$$

where summation is confined to those i 's for which $P(B_i) \neq 0$.⁷

In the limit as n approaches infinity, the probability that the probability given x of whichever element of the partition actually obtains is greater than α is 1. That is the theorem.^{*} What is its significance? According to Savage, "with the observation of an abundance of relevant data, the person is almost certain to become highly convinced of the truth, and it has also been shown that he himself knows this to be the case."⁸ That is a little misleading. The result involves second-order probabilities, but these too, according to personalists, are degrees of belief. So what has been shown seems to be this: In the limit as n approaches infinity an ideally rational Bayesian has degree of belief 1 that an ideally rational Bayesian (with degrees of belief as in the theorem) has degree of belief, given x , greater than α in whichever element of the partition actually obtains. The theorem does not tell us that in the limit any rational Bayesian will assign probability 1 to the true hypothesis and probability 0 to the rest; it only tells us that rational Bayesians are certain that he will. It may reassure those who are already Bayesians, but it is hardly grounds for conversion. Even the reassurance is slim. Mary Hesse⁹ points out, entirely correctly I believe, that the assumptions of the theorem do not seem to apply even approximately in actual scientific contexts. Finally, some of the assumptions of stable estimation theorems can be dispensed with if one assumes instead that all of the initial distributions considered must agree regarding which evidence is relevant to which hypotheses. But there is no evident a priori reason why there should be such agreement.

I think relatively few Bayesians are actually persuaded of the correctness of Bayesian doctrine by Dutch Book arguments, stable estimation theorems, or other a priori arguments. Their frailty is too palpable. I think that the appeal of Bayesian doctrine derives from two other features. First, with only very weak or very natural assumptions about prior probabilities, or none at all, the Bayesian scheme generates principles that seem to accord well with common sense. Thus with minor restrictions one obtains the principle that hypotheses are confirmed by positive instances of them,

^{*} The notation of Savage's theorem has been changed slightly from the way it appears in Glymour's book to make it easier to read. There is an explanation of the theorem in "Bayesian A Priori Arguments" in the commentary on chapter 5.

and, again, one obtains the result that if an event that actually occurs is, on some hypothesis, very unlikely to occur, then that occurrence renders the hypothesis less likely than it would otherwise have been. These principles, and others, can claim something like the authority of common sense, and Bayesian doctrine provides a systematic explication of them. Second, the restrictions placed a priori on rational degrees of belief are so mild, and the device of probability theory at once so precise and so flexible, that Bayesian philosophers of science may reasonably hope to explain the subtleties and vagaries of scientific reasoning and inference by applying their scheme together with plausible assumptions about the distribution of degrees of belief. This seems, for instance, to be Professor Hesse's line of argument. After admitting the insufficiency of the standard arguments for Bayesianism, she sets out to show that the view can account for a host of alleged features of scientific reasoning and inference. My own view is different. particular *inferences* can almost always be brought into accord with the Bayesian scheme by assigning degrees of belief more or less ad hoc, but we learn nothing from this agreement. What we want is an explanation of scientific argument; what the Bayesians give us is a theory of learning, indeed a theory of personal learning. But arguments are more or less impersonal; I make an argument to persuade anyone informed of the premisses, and in doing so I am not reporting any bit of autobiography. To ascribe to me degrees of belief that make my slide from my premisses to my conclusion a plausible one fails to explain anything not only because the ascription may be arbitrary, but also because, even if it is a correct assignment of my degrees of belief, it does not explain why what I am doing is *arguing*—why, that is, what I say should have the least influence on others, or why I might hope that it should. Now, Bayesians might bridge the gap between personal inference and argument in either of two ways. In the first place, one might give arguments in order to change others' beliefs because of the respect they have for his opinion. This is not very plausible, if that were the point of giving arguments, one would not bother with them but would simply state one's opinion. Alternatively, and more hopefully, Bayesians may suggest that we give arguments exactly because there are general principles restricting belief, principles that are widely subscribed to, and in giving arguments we are attempting to show that, supposing our audience has certain beliefs, they must in view of these principles have other beliefs, those we are trying to establish. There is nothing controversial about this suggestion, and I endorse it. What is controversial is that the general principles required for argument can best be understood as conditions restricting prior probabilities in a Bayesian framework. Sometimes they can, perhaps, but I think that when arguments turn on relating evidence to theory, it is very difficult to explicate them in a plausible way within the Bayesian framework. At any rate, it is worth seeing in more detail what the difficulties may be.

There is very little Bayesian literature about the hodgepodge of claims

and notions that are usually canonized as scientific method; very little seems to have been written, from a Bayesian point of view, about what makes a hypothesis *ad hoc*, about what makes one body of evidence more various than another body of evidence, and why we should prefer a variety of evidence, about why, in some circumstances, we should prefer simpler theories, and what it is that we are preferring when we do. And so on. There is little to nothing of this in Carnap, and more recent, and more personalist, statements of the Bayesian position are almost as disappointing. In a lengthy discussion of what he calls "tempered personalism," Abner Shimony¹⁹ discusses only how his version of Bayesianism generalizes and qualifies hypothetico-deductive arguments. (Shimony does discuss simplicity, but only to argue that it is overvalued.) Mary Hesse devotes the later chapters of her book to an attempt to show that certain features of scientific method do result when the Bayesian scheme is supplemented with a postulate that restricts assignments of prior probabilities. Unfortunately, as we shall see, her restrictive principle is incoherent.²⁰

One aspect of the demand for a variety of evidence arises when there is some definite set of alternative hypotheses between which we are trying to decide. In such cases we naturally prefer the body of evidence that will be most helpful in eliminating false competitors. This aspect of variety is an easy and natural one for Bayesians to take account of, and within an account such as Shimony's it is taken care of so directly as hardly to require comment. But there is more to variety. In some situations we have some reason to suspect that if a theory is false, its falsity will show up when evidence of certain kinds is obtained and compared. For example, given the tradition of Aristotelian distinctions, there was some reason to demand both terrestrial and celestial evidence for seventeenth-century theories of motion that subjected all matter to the same dynamical laws. Once again, I see no special reason why this kind of demand for a variety of evidence cannot be fitted into the Bayesian scheme. But there is still more. A complex theory may contain a great many logically independent hypotheses, and particular bodies of evidence may provide grounds for some of those hypotheses but not for others. Surely part of the demand for a variety of evidence, and an important part, derives from a desire to see to it that the various independent parts of our theories are tested. Taking account of this aspect of the demand for a variety of evidence is just taking account of the relevance of evidence to pieces of theory. How Bayesians may do this we shall consider later.

Simplicity is another feature of scientific method for which some Bayesians have attempted to account. There is one aspect of the scientific preference for the simple that seems beyond Bayesian capacities, and that is the disdain for "deocannized" hypotheses, for theories that postulate the operation of a number of properties, determinable only in combination, when a single property would do. Such theories can be generated by taking any ordinary theory and replacing some single quantity, wherever

it occurs in the statement of the theory, by an algebraic combination of new quantities. If the original quantity was not one that occurs in the statement of some body of evidence for the theory, then the new, deoccamized theory will have the same entailment relations with that body of evidence as did the original theory. If the old theory entailed the evidence, so will the new, deoccamized one. Now, it follows from Bayesian principles that if two theories both entail e , then (provided the prior probability of each hypothesis is neither 1 nor 0) if e confirms one of them it confirms the other. How then is the fact (for so I take it to be) that pieces of evidence just don't seem to count for deoccamized theories to be explained? Not by supposing that deoccamized theories have lower prior probabilities than undeoccamized theories, for being "deoccamized" is a feature that a theory has only with respect to a certain body of evidence, and it is not hard to imagine artificially restricted bodies of evidence with respect to which perfectly good theories might count as deoccamized. Having extra wheels is a feature a theory has only in relation to a body of evidence, the only Bayesian relation that appears available and relevant to scientific preference is the likelihood of the evidence on the theory, and unfortunately the likelihood is the same for a theory and for its deoccamized counterparts whenever the theory entails the evidence.

It is common practice in fitting curves to experimental data, in the absence of an established theory relating the quantities measured, to choose the "simplest" curve that will fit the data. Thus linear relations are preferred to polynomial relations of higher degree, and exponential functions of measured quantities are preferred to exponential functions of algebraic combinations of measured quantities, and so on. The problem is to account for this preference. Harold Jeffreys,¹² a Bayesian of sorts, offered an explanation along the following lines. Algebraic and differential equations may be ordered by simplicity; the simpler the hypothetical relation between two or more quantities, the greater is its prior probability. If measurement error has a known probability distribution, we can then compute the likelihood of any set of measurement results given an equation relating the measured quantities. It should be clear, then, that with these priors and likelihoods, ratios of posterior probabilities may be computed from measurement results. Jeffreys constructed a Bayesian significance test for the introduction of higher-degree terms in the equation relating the measured quantities. Roughly, if one's equation fits the data too well, then the equation has too many terms and too many arbitrary parameters, and if the equation does not fit the data well enough, then one has not included enough terms and parameters in the equation. The whole business depends, of course, entirely on the ordering of prior probabilities. In his *Theory of Probability*¹³ Jeffreys proposed that the prior probability of a hypothesis decreases as the number of arbitrary parameters increases, but hypotheses having the same number of arbitrary parameters have the same prior probability. This leads immediately to the conclusion that the prior

probability of every hypothesis is zero. Earlier, Jeffreys proposed a slightly more complex assignment of priors that did not suffer this difficulty. The problem is not really one of finding a way to assign finite probabilities to an infinite number of incompatible hypotheses, for there are plenty of ways to do that. The trouble is that it is just very implausible that scientists typically have their prior degrees of belief distributed according to any plausible simplicity ordering, and still less plausible that they would be rational to do so. I can think of very few simple relations between experimentally determined quantities that have withstood continued investigation, and often simple relations are replaced by relations that are infinitely complex. consider the fate of Kepler's laws. Surely it would be naive for anyone to suppose that a set of newly measured quantities will truly stand in a simple relation, especially in the absence of a well-confirmed theory of the matter. Jeffreys' strategy requires that we proceed in ignorance of our scientific experience, and that can hardly be a rational requirement.

Consider another Bayesian attempt, this one due to Mary Hesse. Hesse puts a "clustering" constraint on prior probabilities: for any positive r , the conjunction of $r + 1$ positive instances of a hypothesis is more probable than a conjunction of r positive instances with one negative instance. This postulate, she claims, will lead us to choose, *ceteris paribus* [other things being equal], the most economical, the simplest, hypotheses compatible with the evidence. Here is the argument:

Consider first evidence consisting of individuals a_1, a_2, \dots, a_n , all of which have properties P and Q . Now consider an individual a_{n+1} with property P . Does a_{n+1} have Q or not? If nothing else is known the clustering postulate will direct us to predict Qa_{n+1} since, *ceteris paribus*, the universe is to be postulated to be as homogeneous as possible consistently with the data. But this is also the prediction that would be made by taking the most economical general law which is both confirmed by the data and of sufficient content to make a prediction about the application of Q to a_{n+1} . For $h = \text{"All } P \text{ are } Q\text{"}$ is certainly more economical than the "gruified" conflicting hypothesis of equal content $h' = \text{"All } x \text{ up to } a_n \text{ that are } P \text{ are } Q, \text{ and all other } x \text{ that are } P \text{ are } \sim Q\text{"}$. [. . .]

It follows in the [case] considered that if a rule is adopted to choose the prediction resulting from the most probable hypothesis on grounds of content, or, in case of a tie in content, the most economical hypothesis of those of equal content, this rule will yield the same predictions as the clustering postulate. [. . .]

Here is the argument applied to curve fitting.

Let f be the assertion that two data points $(x_1, y_1), (x_2, y_2)$ are obtained from an experiment. The two points are consistent with the hypothesis $y = a + bx$, and also of course with an indefinite number of other hypotheses

of the form $y = a_0 + a_1x + a_2x^2 \dots$, where the values of a_0, a_1, \dots are not determined by x_1, y_1, x_2, y_2 . What is the most economical prediction of the y -value of a further point g , where the x -value of g is x_3 ? Clearly it is the prediction which uses only the information already contained in f , that is, the calculable values of a, b rather than a prediction which assigns arbitrary values to the parameters of a higher-order hypothesis. Hence the most economical prediction is about the point $g = (x_3, a + bx_3)$, which is also the prediction given by the "simplest" hypothesis on almost all accounts of the simplicity of curves. Translated into probabilistic language, this is to say that to conform to intuitions about economy we should assign higher initial probability to the assertion that points $(x_1, a + bx_1), (x_2, a + bx_2), (x_3, a + bx_3)$ are satisfied by the experiment, than to that in which the third point is inexpressible in terms of a and b alone. In this formulation economy is a function of finite descriptive lists of points rather than general hypotheses, and the relevant initial probability is that of a universe containing these particular points rather than that of a universe in which the corresponding general law is true.

Description in terms of a minimum number of parameters may therefore be regarded as another aspect of homogeneity or clustering of the universe.

Hesse's clustering postulate applies directly to the curve-fitting case, for her clustering postulate then requires that if two paired values of x and y satisfy the predicate $y = ax + b$, then it is more probable than not that a third pair of values will satisfy the predicate. So the preference for the linear hypothesis in the next instance results from Hesse's clustering postulate and the probability axioms. Unfortunately, with trivial additional assumptions, everything results. For surely if $y = a + bx$ is a legitimate predicate, then so is $y = a_1 + b_1x^2$, for any definite values of a_1 and b_1 . Now Hesse's first two data points can be equally well described by $(x_1, a_1 + b_1x_1^2)$ and $(x_2, a_1 + b_1x_2^2)$, where

$$b_1 = \frac{y_1 - y_2}{x_1^2 - x_2^2} \quad a_1 = y_1 - x_1^2 \left(\frac{y_1 - y_2}{x_1^2 - x_2^2} \right)$$

Hence her first two data points satisfy both the predicate $y = a + bx$ and the predicate $y = a_1 + b_1x^2$. So by the clustering postulate the probability that the third point satisfies the quadratic expression must be greater than one-half, and the probability that the third point satisfies the linear expression must also be greater than one-half, which is impossible.

Another Bayesian account of our preference for simple theories has recently been offered by Roger Rosenkrantz.¹⁵ Suppose that we have some criterion for "goodness of fit" of a hypothesis to data, for example, confidence regions based on the χ^2 distribution for categorical data, or in curve fitting perhaps that the average sum of squared deviations is less than some figure. Where the number of possible outcomes is finite we can compare

the number of such possible outcomes that meet the goodness of fit criterion with the number that do not. This ratio Rosenkrantz calls the "observed sample coverage" of the hypothesis. Where the possible outcomes are infinite, if the region of possible outcomes meeting the goodness-of-fit criterion is always bounded for all relevant hypotheses, we can compare the volumes of such regions for different hypotheses and thus obtain a measure of comparative sample coverage.

It seems plausible enough that the smaller the observed sample coverage of a hypothesis, the more severely it is tested by observing outcomes. Rosenkrantz's first proposal is this: the smaller the observed sample coverage, the simpler the hypothesis. But further, he proves the following for hypotheses about categorical data. If H_1 and H_2 are hypotheses with parameters and H_1 is a special case of H_2 obtained by letting a free parameter in H_2 take its maximum likelihood value, then if we average the likelihood of getting evidence that fits each hypothesis well enough over all the possible parameter values, the average likelihood of H_1 will be greater than the average likelihood of H_2 . The conclusion Rosenkrantz suggests is that the simpler the theory the greater the average likelihood of data that fits it sufficiently well. Hence, even if a simple theory has a lower prior probability than more complex theories, because the average likelihood is higher for the simple theory, its posterior probability will increase more rapidly than that of more complex theories. When sufficient evidence has accumulated the simple theory will be preferred. Rosenkrantz proposes to identify average likelihood with support.

Rosenkrantz's approach has many virtues, I shall concentrate on its vices. First, observed sample coverage does not correlate neatly with simplicity. If H is a hypothesis, T another utterly irrelevant to H and to the phenomena about which H makes predictions, then $H \& T$ will have the same observed sample coverage as does H . Further, if H^* is a deoccamization of H , then H^* and H will have the same observed sample coverage. Second, Rosenkrantz's theorem does not establish nearly enough. It does not establish, for example, that in curve fitting the average likelihood of a linear hypothesis is greater than the average likelihood of a quadratic or higher-degree hypothesis. We cannot explicate support in terms of average likelihood unless we are willing to allow that evidence supports a deoccamized hypothesis as much as undeoccamized ones, and a hypothesis with tacked-on parts as much as one without such superfluous parts.

Finally, we come to the question of the relevance of evidence to theory. When does a piece of evidence confirm a hypothesis according to the Bayesian scheme of things? The natural answer is that it does so when the posterior probability of the hypothesis is greater than its prior probability, that is, if the conditional probability of the hypothesis on the evidence is greater than the probability of the hypothesis. That is what the condition of positive relevance requires, and that condition is the one most commonly advanced by philosophical Bayesians. The picture is a kine-

matic one: a Bayesian agent moves along in time having at each moment a coherent set of degrees of belief; at discrete intervals he learns new facts and each time he learns a new fact, e , he revises his degrees of belief by conditionalizing on e . The discovery that e is the case has confirmed those hypotheses whose probability after the discovery is higher than their probability before. For several reasons, I think this account is unsatisfactory, moreover, I doubt that its difficulties are remediable without considerable changes in the theory.

The first difficulty is a familiar one. Let us suppose that we can divide the consequences of a theory into sentences consisting of reports of actual or possible observations, and simple generalizations of such observations, on the one hand, and on the other hand, sentences that are theoretical. Then the collection of "observational" consequences of the theory will always be at least as probable as the theory itself, generally, the theory will be less probable than its observational consequences. A theory is never any better established than is the collection of its observational consequences. Why, then, should we entertain theories at all? On the probabilist view, it seems, they are a gratuitous risk. The natural answer is that theories have some special function that their collection of observational consequences cannot serve; the function most frequently suggested is explanation—theories explain, their collection of observational consequences do not. But however sage this suggestion may be, it only makes more vivid the difficulty of the Bayesian way of seeing things. For whatever explanatory power may be, we should certainly expect that goodness of explanation will go hand in hand with warrant for belief, yet if theories explain and their observational consequences do not, the Bayesian must deny the linkage. The difficulty has to do both with the assumption that rational degrees of belief are generated by probability measures and with the Bayesian account of evidential relevance. Making degrees of belief probability measures in the Bayesian way already guarantees that a theory can be no more credible than any collection of its consequences. The Bayesian account of confirmation makes it impossible for a piece of evidence to give us more total credence in a theory than in its observational consequences. The Bayesian way of setting things up is a natural one, but it is not inevitable, and wherever a distinction between theory and evidence is plausible, it leads to trouble.

A second difficulty has to do with how praise and blame are distributed among the hypotheses of a theory. Recall the case of Kepler's laws It seems that observations of a single planet (and, of course, the sun) might provide evidence for or against Kepler's first law (all planets move on ellipses) and for or against Kepler's second law (all planets move according to the area rule), but no observations of a single planet would constitute evidence for or against Kepler's third law (for any two planets, the ratio of their periods equals the $\frac{3}{2}$ power of the ratio of their distances) Hypothetico-deductive accounts of confirmation have great dif-

difficulty explaining this elementary judgment. Can the Bayesians do any better? One thing that Bayesians can say (and some have said) is that our degrees of belief are distributed—and historically were distributed—so that conditionalizing on evidence about one planet may change our degrees of belief in the first and second laws, but not our degree of belief in the third law.¹⁶ I don't see that this is an explanation for our intuition at all; on the contrary, it seems merely to restate (with some additional claims) what it is that we want to be explained. Are there any reasons why people had their degrees of belief so distributed? If their beliefs had been different, would it have been equally rational for them to view observations of Mars as a test of the third law, but not of the first? It seems to me that we never succeed in explaining a widely shared judgment about the relevance or irrelevance of some piece of evidence merely by asserting that degrees of belief happened to be so distributed as to generate those judgments according to the Bayesian scheme. Bayesians may instead try to explain the case by appeal to some structural difference among the hypotheses, the only gadget that appears to be available is the likelihood of the evidence about a single planet on various combinations of hypotheses. If it is supposed that the observations are such that Kepler's first and second laws entail their description, but Kepler's third law does not, then it follows that the likelihood of the evidence on the first and second laws—that is, the conditional probability of the evidence given those hypotheses—is unity, but the likelihood of the evidence on the third law may be less than unity. But any attempt to found an account of the case on these facts alone is simply an attempt at a hypothetico-deductive account. The problem is reduced to one already unsolved. What is needed to provide a genuine Bayesian explanation of the case in question (as well as of many others that could be adduced) is a *general* principle restricting conditional probabilities and having the effect that the distinctions about the bearing of evidence that have been noted here do result. Presumably any such principles will have to make use of relations of content or structure between evidence and hypothesis. The case does nothing to establish that no such principles exist; it does, I believe, make it plain that without them the Bayesian scheme does not *explain* even very elementary features of the bearing of evidence on theory.

A third difficulty has to do with Bayesian kinematics. Scientists commonly argue for their theories from evidence known long before the theories were introduced. Copernicus argued for his theory using observations made over the course of millenia, not on the basis of any startling new predictions derived from the theory, and presumably it was on the basis of such arguments that he won the adherence of his early disciples. Newton argued for universal gravitation using Kepler's second and third laws, established before the *Principia* was published. The argument that Einstein gave in 1915 for his gravitational field equations was that they explained the anomalous advance of the perihelion of Mercury, established

more than half a century earlier. Other physicists found the argument enormously forceful, and it is a fair conjecture that without it the British would not have mounted the famous eclipse expedition of 1919. Old evidence can in fact confirm new theory, but according to Bayesian kinematics it cannot. For let us suppose that evidence e is known before theory T is introduced at time t . Because e is known at t , $\text{Prob}_t(e) = 1$. Further, because $\text{Prob}_t(e) = 1$, the likelihood of e given T , $\text{Prob}_t(e|T)$, is also 1. We then have

$$\text{Prob}_t(T|e) = \frac{\text{Prob}_t(T) \times \text{Prob}_t(e|T)}{\text{Prob}_t(e)} = \text{Prob}_t(T)$$

The conditional probability of T on e is therefore the same as the prior probability of T : e cannot constitute evidence for T in virtue of the positive relevance condition nor in virtue of the likelihood of e on T . None of the Bayesian mechanisms apply, and if we are strictly limited to them, we have the absurdity that old evidence cannot confirm new theory. The result is fairly stable. If the probability of e is very high but not unity, $\text{Prob}_t(e|T)$ will still be unity if T entails e , and so $\text{Prob}_t(T|e)$ will be very close to $\text{Prob}_t(T)$. How might Bayesians deal with the old evidence/new theory problem?¹⁷ Red herrings abound: the prior probability of the evidence, Bayesians may object, is not really unity; when the evidence is stated as measured or observed values, the theory does not really entail that those exact values obtain; an ideal Bayesian would never suffer the embarrassment of a novel theory. None of these replies will do: the acceptance of old evidence may make the degree of belief in it as close to unity as our degree of belief in some bit of evidence ever is; although the exact measured value (of, for example, the perihelion advance) may not be entailed by the theory and known initial conditions, that the value of the measured quantity lies in a certain interval may very well be entailed, and that is what is believed anyway; and, finally, it is beside the point that an ideal Bayesian would never face a novel theory, for the idea of Bayesian confirmation theory is to explain scientific inference and argument by means of the assumption that good scientists are, about science at least, approximately ideal Bayesians, and we have before us a feature of scientific argument that seems incompatible with that assumption.

A natural line of defense lies through the introduction of counterfactual degrees of belief. When using Bayes' rule to determine the posterior probability of a new theory on old evidence, one ought not to use one's actual degree of belief in the old evidence, which is unity or nearly so.

* In his book, Glymour writes the conditional probability of e given T as $\text{Prob}(e, T)$ and, later, as $P(e, T)$. We have substituted the slash notation for conditional probabilities used by the other authors in this chapter.

one ought instead to use the degree of belief one would have had in e if . . . The problem is to fill in the blanks in such a way that it is both plausible that we have the needed counterfactual degrees of belief and that they do serve to determine how old evidence bears on new theory. I tend to doubt that there is such a completion. We cannot merely throw e and whatever entails e out of the body of accepted beliefs; we need some rule for determining a counterfactual degree of belief in e and a counterfactual likelihood of e on T . To simplify, let us suppose that T does logically entail e , so that the likelihood is fixed.

If one flips a coin three times and it turns up heads twice and tails once, in using this evidence to confirm hypotheses (e.g., of the fairness of the coin) one does not take the probability of two heads and one tail to be what it is after the flipping—namely, unity—but what it was before the flipping. In this case there is an immediate and natural counterfactual degree of belief that is used in conditionalizing by Bayes' rule. The trouble with the scientific cases is that no such immediate and natural alternative distribution of degree of belief is available. Consider someone trying, in a Bayesian way, to determine in 1915 how much Einstein's derivation of the perihelion advance confirmed general relativity. There is no single event, like the coin flipping, that makes the perihelion anomaly virtually certain. Rather, Leverrier first computed the anomaly in the middle of the nineteenth century; Simon Newcomb calculated it again around 1890 using Leverrier's method but new values for planetary masses, and obtained a substantially higher value than had Leverrier. Both Newcomb and Leverrier had, in their calculations, approximated an infinite series by its first terms without any proof of convergence, thus leaving open the possibility that the entire anomaly was the result of a mathematical error. In 1912 Eric Doolittle calculated the anomaly by a wholly different method, free of any such assumption, and obtained virtually the same value as had Newcomb.¹⁸ For actual historical cases, unlike the coin-flipping case, there is no single counterfactual degree of belief in the evidence ready to hand, for belief in the evidence sentence may have grown gradually—in some cases it may have even waxed, waned, and waxed again. So the old evidence/new theory problem cannot be assimilated to coin flipping.

The suggestion that what is required is a counterfactual degree of belief is tempting nonetheless; but there are other problems with it besides the absence of any unique historical degree of belief. A chief one is that various ways of manufacturing counterfactual degrees of belief in the evidence threaten us with incoherence. One suggestion, for example, is the following, used implicitly by some Bayesian writers. At about the time T is introduced, there will be a number of alternative competing theories available; call them T_1, T_2, \dots, T_k , and suppose that they are mutually exclusive of T and of each other. Then $P(e)$ is equal to

$$P(T_1) \times P(e|T_1) + P(T_2) \times P(e|T_2) + \dots + P(T_k) \times P(e|T_k) \\ + P(\sim(T_1 \vee \dots \vee T_k)) \times P(e|\sim(T_1 \vee \dots \vee T_k))$$

and we may try to use this formula to evaluate the counterfactual degree of belief in e . The problem is with the last term. Of course, one could suggest that this term just be ignored when evaluating $P(e)$, but it is difficult to see within a Bayesian framework any rationale at all for doing so. For if one does ignore this term, then the collection of prior probabilities used to evaluate the posterior probability of T will not be coherent unless either the likelihood of e on T is zero or the prior probability of T is zero. One could remedy this objection by replacing the last term by

$$P(T) \times P(e|T)$$

but this will not do either, for if one's degree of belief in

$$P(T_1 \vee T_2 \vee \dots \vee T_k \vee T)$$

is not unity, then the set of prior degrees of belief will still be incoherent. Moreover, not only will it be the case that if the actual degree of belief in e is replaced by a counterfactual degree of belief in e according to either of these proposals then the resulting set of priors will be incoherent, it will further be the case that if we conditionalize on e the resulting conditional probabilities will be incoherent. For example, if we simply delete the last term, one readily calculates that

$$P(T_1 \vee \dots \vee T_k|e) = \frac{P(T_1 \vee \dots \vee T_k) \times P(e|T_1 \vee \dots \vee T_k)}{P(e|T_1 \vee \dots \vee T_k) \times P(T_1 \vee \dots \vee T_k)} = 1$$

and further that

$$P(T|e) = \frac{P(T) \times P(e|T)}{P(e|T_1 \vee \dots \vee T_k) \times P(T_1 \vee \dots \vee T_k)}$$

But because T is supposed inconsistent with $T_1 \vee \dots \vee T_k$ and $P(T|e)$ is not zero, this is incoherent.

Let us return to the proposal that when new theory confronts old evidence we should look backwards to the time when the old evidence e had not yet been established and use for the prior probability of e whatever degree of belief we would have had at that time. We cannot just stick in such a counterfactual value for the prior probability of e and change nothing else without, as before, often making both prior and conditionalized probabilities incoherent. If we give all of our sentences the degree of belief they would have had in the relevant historical period (supposing we some-

how know what period that is) and then conditionalize on e , incoherency presumably will not arise, but it is not at all clear how to combine the resulting completely counterfactual conditional probabilities with our actual degrees of belief. It does seem to me that the following rather elaborate procedure will work when a new theory is introduced. Starting with your actual degree of belief function P , consider the degree of belief you would have had in e in the relevant historical period, call it $H(e)$. Now change P by regarding $H(e)$ as an arbitrary change in degree of belief in e and using Richard Jeffrey's¹⁹ rule

$$P'(S) = H(e) \times P(S|e) + (1 - H(e)) \times P(S|\sim e)$$

Jeffrey's rule guarantees that P' is a probability function. Finally, conditionalize on e

$$P''(S) = P'(S|e)$$

and let P'' be your new actual degree of belief function. (Alternatively, P'' can be formed by using Jeffrey's rule a second time)

There remain a number of objections to the historical proposal. It is not obvious that there are, for each of us, degrees of belief we personally would have had in some historical period. It is not at all clear which historical period is the relevant one. Suppose, for example, that the gravitational deflection of starlight had been determined experimentally around 1900, well before the introduction of general relativity.²⁰ In trying to assess the confirmation of general relativity, how far back in time should a twentieth-century physicist go under this supposition? If only to the nineteenth, then if he would have shared the theoretical prejudices of the period, gravitational deflection of light would have seemed quite probable. Where ought he to stop and why? But laying aside these difficulties, it is implausible indeed that such a historical Bayesianism, however intriguing a proposal, is an accurate account of the principles by which scientific judgments of confirmation are made. For if it were, then we should have to condemn a great mass of scientific judgments on the grounds that those making them had not studied the history of science with sufficient closeness to make a judgment as to what their degrees of belief would have been in relevant historical periods. Combined with the delicacy that is required to make counterfactual degrees of belief fit coherently with actual ones, these considerations make me doubt that we should look to counterfactual degrees of belief for a plausible Bayesian account of how old evidence bears on new theory.

Finally, consider a quite different Bayesian response to the old evidence/new theory problem. Whereas the ideal Bayesian agent is a perfect logician, none of us are, and there are always consequences of our hypotheses that we do not know to be consequences. In the situation in

which old evidence is taken to confirm a new theory, it may be argued that there is *something* new that is learned, and typically, what is learned is that the old evidence is entailed by the new theory. Some old anomalous result is lying about, and it is not this old result that confirms a new theory, but rather the new discovery that the new theory entails (and thus explains) the old anomaly. If we suppose that semirational agents have degrees of belief about the entailment relations among sentences in their language and that

$$P(h \vdash e) = 1 \text{ implies } P(e|h) = 1$$

this makes a certain amount of sense.* We imagine the semirational Bayesian changing his degree of belief in hypothesis h in light of his new discovery that h entails e by moving from his prior degree of belief in h to his conditional degree of belief in h given that e , that $h \vdash e$, and whatever background beliefs there may be. Old evidence can, in this vicarious way, confirm a new theory, then, provided that

$$P(h|b \ \& \ e \ \& \ (h \vdash e)) > P(h|b \ \& \ e)$$

Now, in a sense, I believe this solution to the old evidence/new theory problem to be the correct one; what matters is the discovery of a certain logical or structural connection between a piece of evidence and a piece of theory, and it is in virtue of that connection that the evidence, if believed to be true, is thought to be evidence for the bit of theory. What I do not believe is that the relation that matters is simply the entailment relation between the theory, on the one hand, and the evidence on the other. The reasons that the relation cannot be simply that of entailment are exactly the reasons why the hypothetico-deductive account . . . is inaccurate, but the suggestion is at least correct in sensing that our judgment of the relevance of evidence to theory depends on the perception of a structural connection between the two, and that degree of belief is, at best, epiphenomenal. In the determination of the bearing of evidence on theory there seem to be mechanisms and strategems that have no apparent connection with degrees of belief, which are shared alike by people advocating different theories. Save for the most radical innovations, scientists seem to be in close agreement regarding what would or would not be evidence relevant to a novel theory; claims as to the relevance to some hypothesis of some observation or experiment are frequently buttressed by detailed calculations and arguments. All of these features of the determination of evidential relevance suggest that that relation depends somehow on structural, objective features connecting statements of evidence and

* The turnstile symbol, \vdash , denotes entailment. One statement entails another when the second is logically deducible or derivable from the first.

statements of theory. But if that is correct, what is really important and really interesting is what these structural features may be. The condition of positive relevance, even if it were correct, would simply be the least interesting part of what makes evidence relevant to theory.

None of these arguments is decisive against the Bayesian scheme of things, nor should they be, for in important respects that scheme is undoubtedly correct. But taken together, I think they do at least strongly suggest that there must be relations between evidence and hypotheses that are important to scientific argument and to confirmation but to which the Bayesian scheme has not yet penetrated.

■ | Notes

1. "Who cares whether a pig-farmer is a Bayesian?"—R. C. Jeffrey

[In the preface to *Theory and Evidence*, from which this reading is taken, Glymour explains that most of the first draft of his book was written during a year's leave from Princeton when he lived with his family in a one-room shack in rural Virginia. He did raise some pigs. In a letter to the editors of the present volume he relates, "I spent a lot of time there thinking about Bayesian stuff, and when I returned, each time I would meet Dick Jeffrey on campus, I would put to him one or another of the puzzles I had found. On one occasion, perhaps growing a little exasperated at the game, or perhaps because the puzzle was more difficult, he gave out the infamous line you ask about." Just for the record, Richard Jeffrey (a Bayesian) would like it to be known that some of his best friends are pig farmers.]

2. R. Carnap, *The Logical Foundations of Probability*, Chicago: University of Chicago Press, 1950.

3. See Hilary Putnam, "Probability and Confirmation," in S. Morgenbesser, *Philosophy of Science Today*, New York: Basic Books, 1967.

4. A third view, that probabilities are to be understood exclusively as frequencies, has been most ably defended by Wesley Salmon. See his *Foundations of Scientific Inference*, Pittsburgh: University of Pittsburgh Press, 1969.

5. More detailed accounts of means for determining degrees of belief may be found in R. Jeffrey, *The Logic of Decision*, New York: McGraw-Hill, 1965. It is a curious fact that the procedures that Bayesians use for determining subjective degrees of belief empirically are an instance of the general strategy [I call bootstrapping] described in Chapter V [of my *Theory and Evidence*]. Indeed, the strategy typically used to determine whether or not actual people behave as rational Bayesians involves the bootstrap strategy described in that chapter.

6. For further criticisms of the Dutch Book argument see H. Kyburg, "Subjective Probability: Criticisms, Reflections, and Problems," *Journal of Philosophical Logic* 7, 1978, pp. 157–180.

7. L. Savage, *The Foundations of Statistics*, New York: Dover, 1972, p. 49.

8. *Ibid.*, p. 50.

9. See M. Hesse, *The Structure of Scientific Inference*, Berkeley: University of California Press, 1974, pp. 117–119.

10. See A. Shimony, "Scientific Inference," in R. Colodny, *The Nature and Function of Scientific Theories*, Pittsburgh: University of Pittsburgh Press, 1970.

11. Moreover, I believe that much of her discussion of methodological principles has only the loosest relation to Bayesian principles.

12. See I. Jeffreys, *Scientific Inference*, Cambridge: Cambridge University Press, 1973.

13. I. Jeffreys, *Theory of Probability*, Oxford: Clarendon Press, 1967.

14. Hesse, *The Structure of Scientific Inference*, pp. 230–232.

15. See R. Rosenkrantz, "Simplicity," in W. Harper and C. Hooker, *Foundations and Philosophy of Statistical Inference*, Boston: Reidel, 1976.

16. This is the account suggested by P. Horwich in "An Appraisal of Glymour's Confirmation Theory," *Journal of Philosophy* 75, 1978, pp. 98–113.

17. All of the defenses sketched below were suggested to me by one or another philosopher sympathetic to the Bayesian view; I have not attributed the arguments to anyone for fear of misrepresenting them. Nonetheless, I thank Jon Dorling, Paul Teller, Daniel Garber, Ian Hacking, Patrick Suppes, Richard Jeffrey, and Roger Rosenkrantz for valuable discussions and correspondence on the point at issue.

18. The actual history is still more complicated. Newcomb and Doolittle obtained values for the anomaly differing by about two seconds of arc per century. Early in the 1920s, Grossmann discovered that Newcomb had made an error in calculation of about that magnitude.

19. See Jeffrey, *The Logic of Decision*.

20. Around 1900 is fanciful, before general relativity is not. In 1914 E. Freundlich mounted an expedition to Russia to photograph the eclipse of that year in order to determine the gravitational deflection of starlight. At that time, Einstein had predicted an angular deflection for light passing near the limb of the sun that was equal in value to that derived from Newtonian principles by Soldner in 1801. Einstein did not obtain the field equations that imply a value for the deflection equal to twice the Newtonian value until late in 1915. Freundlich was caught in Russia by the outbreak of World War I and was interned there. Measurement of the deflection had to wait until 1919.

Wittgensteinian Bayesianism

Belief is not an all-or-nothing matter. Rather, there are various *degrees* of conviction which may be represented by numbers between zero and one. Were we ideally rational, our full beliefs (of degree one) would comply with the laws of deductive logic; they would be consistent and closed under logical implication. And similarly, our *degrees* of belief should conform to the probability calculus.¹ This enrichment of epistemology—provided by the addition of degrees of belief and an appreciation of their probabilistic ‘logic’—fosters progress with respect to many problems in the philosophy of science.

These statements form the core of a program, which I will call “therapeutic Bayesianism,” whose primary goal is the solution of various puzzles and paradoxes that come from reflecting on scientific method. Its creed is that many of these problems are the product of oversimplification, and that the above-mentioned elementary probabilistic model of degrees of belief often contains just the right balance of accuracy and simplicity to enable us to command a clear view of the issues and see where we were going wrong.² This somewhat Wittgensteinian goal and creed distinguishes therapeutic Bayesianism from more systematic enterprises in which probabilistic degrees of belief play a prominent role: for example, Bayesian decision theory, Bayesian statistics, Bayesian psychology, Bayesian semantics, and Bayesian history of science. It is especially important to appreciate the difference between the problem-solving orientation of therapeutic Bayesianism—that of exploiting a simple, idealized model in order to help illuminate notorious philosophical perplexities—and the quite distinct project of providing a perfectly true and complete (descriptive or normative) *theory* of scientific practice. The latter task might well involve the postu-

FROM P. A. French, T. E. Uehling, Jr., and H. K. Wettstein, eds., *Midwest Studies in Philosophy*, vol. 18 (Notre Dame, Ind.: University of Notre Dame Press, 1993), 62–77.

lation of belief-gradations, and might also be done in the name of philosophy of science. However, its aims are quite different; and one must beware of judging one project by adequacy conditions appropriate to the other.³

Therapeutic Bayesianism is not self-evidently beneficial, but it does have some *prima facie* plausibility. Moreover, this plausibility is enhanced by substantial accomplishments, and, as we shall see, a great deal of the criticism it has received is misdirected—commonly for the reason just indicated. In this paper I would like to try to make a case for the program by discussing it from three, progressively abstract, points of view: substantial, foundational, and metaphilosophical. More specifically, there will follow sections on. (I) “The fruitfulness of therapeutic Bayesianism,” in which I will sketch treatments of the ‘raven’ paradox and the question of diverse data and mention various other applications; (II) “Probabilistic foundations,” in which the propriety of certain idealizations will be defended—particularly the representation of belief by numbers, the adoption of probabilistic canons of reason governing such beliefs, the definition of confirmation as increase in rational degree of belief, and the idea that induction may be codified in a confirmation function; and (III) “Misplaced scientism,” in which I criticize a metaphilosophical perspective that does not properly distinguish science from the philosophy of science, and which overvalues the use of symbolic apparatus. Along the way, I shall respond to some criticisms of therapeutic Bayesianism that have recently been advanced.

I | The Fruitfulness of Therapeutic Bayesianism

A good illustration of therapeutic Bayesianism at work is its way of treating the notorious ‘raven paradox’. It is plausible to suppose that any hypothesis of the form ‘All *F*s are *G*’ would be supported by the observation of an *F* that is also *G*. But if this is generally true, then the discovery of a non-black non-raven (e.g., a white shoe) confirms that all non-black things are non-ravens; and thereby confirms the logically equivalent hypothesis, ‘All ravens are black’—a seemingly bizarre conclusion. This is ‘the paradox of confirmation’. The Bayesian approach to this problem is to argue that observing a known raven to be black will *substantially* confirm “All ravens are black,” whereas observing that a known non-black thing is not a raven will confirm it only *negligibly*—the difference being explained, roughly speaking, by the fact that, given our background beliefs about the chances of coming across ravens and black things, the first of these observations is more surprising, more of a test of the hypothesis, and therefore more evidentially powerful, than the second. Thus, the paradoxical flavor of our conclusion comes from the not unnatural confusion of negligible support

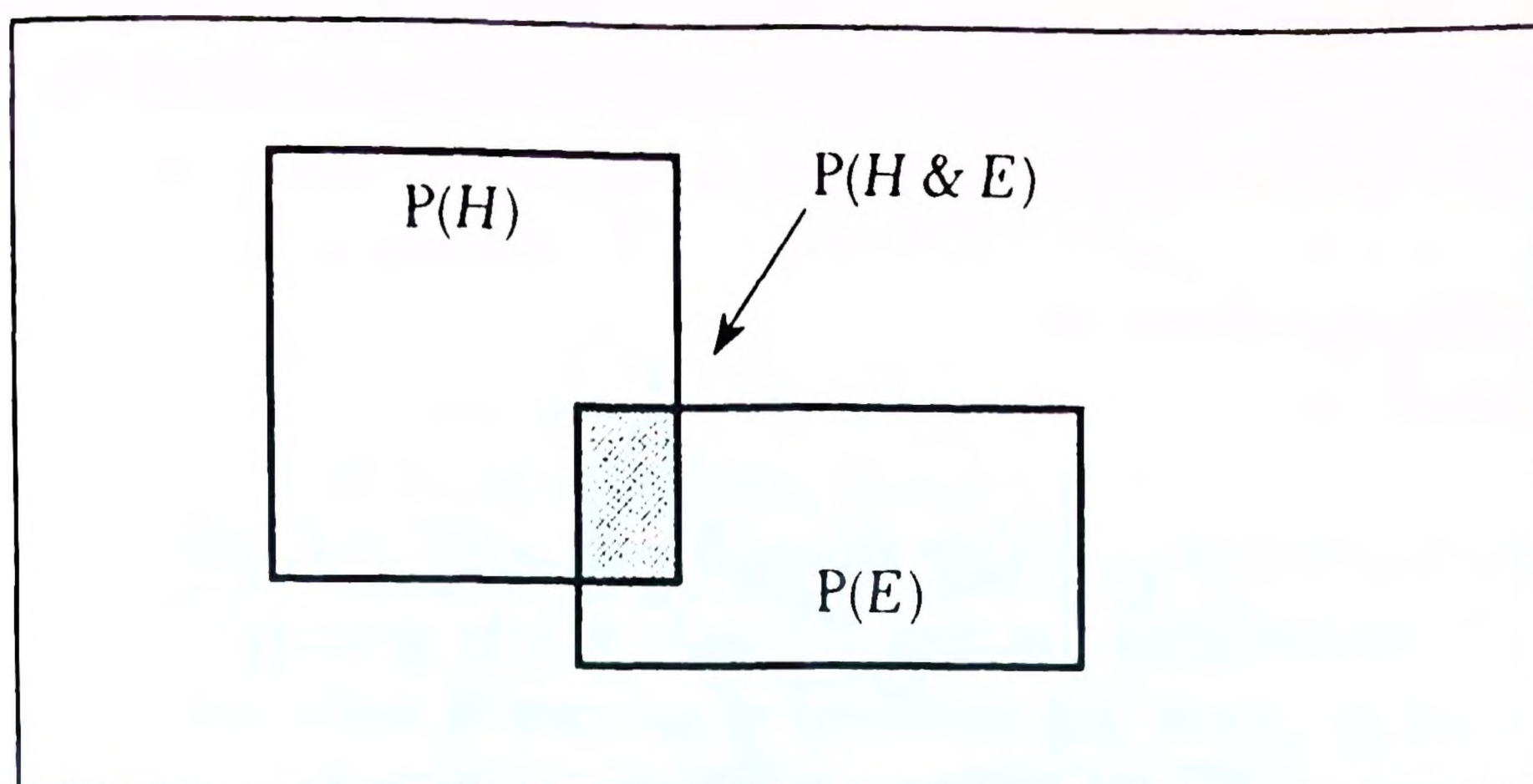


Figure 1

with no support at all—a confusion sustained by inattention to degrees of belief and their bearing on confirmation.

A formal version of this analysis proceeds from the following premises:

- a That the amount of support for hypothesis H provided by evidence E is the factor by which the rational degree of belief in H is enhanced by the discovery of E —which is indicated by the ratio of subjective probabilities, $P(H/E)/P(H)$, for a rational person.
- b That a rational person's degrees of belief will ideally conform to the probability calculus; and, in particular, will obey Bayes's Theorem:

$$\frac{P(H/E)}{P(H)} = \frac{P(E/H)}{P(E)}$$

(To appreciate the intuitive plausibility of this theorem, note that it derives from the fact that the conditional probability of H given E is equal to the probability of the conjunction of H and E , divided by the probability of E : i.e., $P(H/E) = P(H \& E)/P(E)$. See figure 1. Therefore, since $P(H \& E) = P(E \& H)$, we obtain $P(H/E)P(E) = P(E/H)P(H)$, and hence Bayes's Theorem).

- c That our degree of belief (prior to the investigation, and given the known scarcity of ravens) that a randomly selected non-black thing would turn out to be a non-raven is high.
- d That our prior degree of belief (prior to the investigation, and given the known abundance of non-black things) that a randomly selected raven would turn out not to be black is substantial.

Now let us compare the support for the hypothesis, H , that all ravens are black, provided, first, by the discovery concerning a known raven that

it is black (which is symbolized as $R^* B$), and, second, by the discovery that a known non-black thing is not a raven ($-B^* -R$). Applying premise (a) and then (b), we find:

$$\text{Support provided by } (R^* B) = \frac{P(H/R^* B)}{P(H)} = \frac{P(R^* B/H)}{P(R^* B)}$$

$$\text{Support provided by } (-B^* -R) = \frac{P(H/-B^* -R)}{P(H)} = \frac{P(-B^* -R/H)}{P(-B^* -R)}$$

But our hypothesis *entails* that any known raven would be black and any known non-black thing would not be a raven; therefore, $P(R^* B/H) = 1$ and $P(-B^* -R/H) = 1$. Therefore

$$\text{Support from raven found to be black} = \frac{P(H/R^* B)}{P(H)} = \frac{1}{P(R^* B)} = \frac{1/\text{prior degree of belief that a known raven would be black}}$$

$$\text{Support from non-black thing found not to be a raven} = \frac{P(H/-B^* -R)}{P(H)} = \frac{1}{P(-B^* -R)} = \frac{1/\text{prior degree of belief that a known non-black thing would not be a raven}}$$

Now one may assume (premise (c)) that a normal investigator of the hypothesis has prior background knowledge about the rough distribution of ravens and black things in his vicinity, and that this will lead him to expect that there is a very good chance that a randomly selected non-black thing will turn out not to be a raven. Thus $P(-B^* -R)$ is very nearly 1; and the amount of support for the hypothesis provided by observing that a non-black thing is not a raven is very little.

On the other hand, we would expect the background of investigation to dictate, in addition, (premise (d)) that the likelihood of a randomly selected raven being black is not especially high. After all, as far as we know at the outset of the research, there are many colors that the raven could perfectly well have. Thus $P(R^* B)$ is a good deal less than one. Therefore, the amount of support provided by observing that a known raven is black is substantial.

One might object to this reasoning that the final assumption is false, since the *objective* chances of finding that a raven is black are actually extremely high. However, this objection is based on a slip which is easy to identify. It confuses "probability" in the sense of *subjective degree of belief* and "probability" in the sense of *relative frequency*. All the probabilities mentioned in the argument are rational subjective probabilities, and it is under that construal that we may reasonably assume that $P(R^* B)$ is not near to 1. The feeling that this assumption is wrong derives

from incorrectly reading $P(R^* B)$ as a relative frequency assertion. In that sense, since in fact almost all ravens are black, the probability that a randomly selected raven will be black is indeed very great. But this fact has no bearing on the argument.⁴

A similar objection is to deny that there could be any difference in evidential import between identifying a known raven as black and identifying a known black thing as a raven. Howson and Urbach,⁵ for example, maintain that the only difference between these two data is the time order in which the elements of the observed fact are established. They think that in each case what is eventually known is the same, so there can be no variation in confirmation power between the two discoveries. Imagine, however, that an ornithologist instructs her assistant to go and find a black raven and bring it back to the lab for inspection. Surely, that inspection would count for nothing. And there is no paradox here, even though we might loosely speak of 'seeing a black raven' in all three cases. For a more precise characterization of the evidence shows that what is discovered in each case is not really the same. That a randomly selected raven turns out to be black, that a randomly selected black thing turns out to be a raven and that a randomly selected black raven turns out to be a black raven, are very different pieces of information, and it should not be surprising that they confirm our hypothesis to different degrees.

Therapeutic Bayesianism handles other issues in the philosophy of science similarly, putting a lot of weight on premises (a) and (b). By combining the idea of confirmation as enhancement of rational degree of belief, with the principle that rational degrees of belief should satisfy the probability calculus, we get a way of treating those problems that hinge upon considerations having to do with *degree of support*. Therefore the method has a wide scope. In particular, one can expect to shed light on why 'surprising' predictions have relatively great confirmation power, what is wrong with ad hoc hypotheses, whether prediction has more evidential value than mere accommodation of data, why a broad spectrum of facts can confirm a theory more than a narrow data set, why we base our judgments on as much data as possible, how statistical hypotheses can be testable despite their unfalsifiability, what is peculiar about 'grue-like' hypotheses, and various other problems.*

These issues are unified by their involvement with the notion of 'varying evidential quality'; and this is why traditional epistemology, with its fixation on all-or-nothing belief, is not able to resolve them. It is only to be expected that the introduction of degrees of belief, together with an understanding of the rational constraints to which they are subject, would open the way to progress. Of course, there is not the space here to fully substantiate this thesis by describing all these applications of therapeu-

* For a discussion of grue-like hypotheses, see "Goodman's Gruesome New Riddle of Induction" in the commentary on chapter 3.

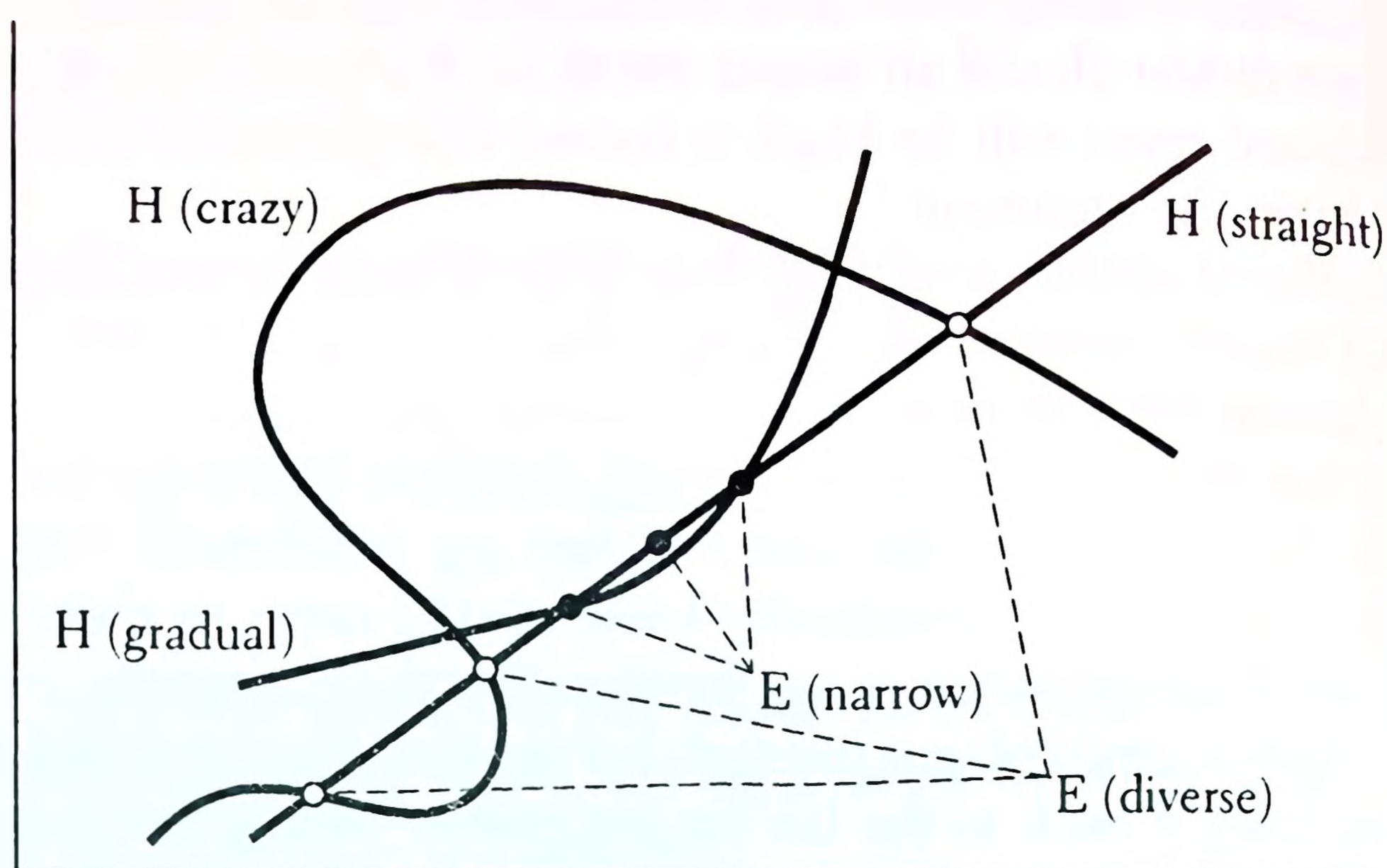


Figure 2

tic Bayesianism. Let me, however, give one further illustration of the approach.

How is it that a broad spectrum of different kinds of fact, when entailed by a hypothesis, will confirm it to a greater degree than a uniform, repetitive set of data? It is natural to answer as follows. To the extent that our observations cover a broad range of phenomena, they are capable of falsifying a large number of alternative hypotheses, which then bequeath substantial credibility to those hypotheses that survive. Now, this solution does not quite work. For a narrow data set can preclude just as *many* hypotheses as a diverse data set. Nevertheless, we can repair the solution by noting that there is a significant difference in the *kinds* of hypotheses that are excluded by the two sets of facts. We should notice that the diverse data tend to exclude more of the *simple* hypotheses than do the narrow data. Given a representation of simplicity in terms of high prior probability,⁶ this suggests that diverse data tend to rule out more high-probability alternatives than narrow data. But if so, then a hypothesis that survives relatively diverse observations becomes more probable than one that is left in the running by a narrow set of data. (See figure 2).

In particular, the data points E(narrow) exclude (given experimental error) just as many alternatives to the line H(straight) as does E(diverse). Nonetheless E(diverse) confirms H(straight) more strongly than E(narrow) does, because E(diverse) is better than E(narrow) at excluding simple alternatives to H(straight)—for example, gradual curves—which have an initially high probability. Thus H(gradual) is ruled out by E(diverse) but not by E(narrow). On the other hand, the sort of hypothesis, like H(crazy), prohibited by E(narrow) yet not by E(diverse), is not very probable anyway; so excluding it does not greatly benefit those hypotheses that survive. Thus,

with the help of a probabilistic representation of simplicity, we can begin to account for our methodological intuitions concerning diverse data.⁷

II | Probabilistic Foundations

In the last section I have tried to indicate something of the fruitfulness of therapeutic Bayesianism. Let me now consider various foundational questions that might be thought to cast doubt on the project:

- 1 Do people actually have numerical degrees of belief?
- 2 If so, can it be shown that *rational* degrees of belief conform to the probability calculus?
- 3 Is it correct to identify degree of confirmation with rational enhancement of subjective probability?
- 4 Are there objective facts of confirmation?
- 5 Does reason require *merely* that one's beliefs conform to the probability calculus? Or is it the case (as Carnap thought) that a rational system of beliefs is subject to several further constraints?
- 6 If further constraints are needed, then what are they?

On the first question, perhaps we should be agnostic. The successes of therapeutic Bayesianism will reinforce the evident fact that its basic principles are at least *roughly* correct. Thus we know that there are belief gradations of some sort, that there are rational constraints governing them (prohibiting, for example, a high degree of confidence in two contradictory propositions), and that confirmation is not wholly unrelated to increasing belief. Moreover, the Bayesian representation of these ideas has a great deal of plausibility. Consider a spectrum of situations in which we know that the propensities (as manifested by relative frequencies) of certain events are x_1, x_2, \dots, x_N . For each such case there is a corresponding epistemic attitude—a degree of confidence—that the next trial will produce an event of the designated type. Presumably the appropriate attitude will vary with the relative frequency. Specifically, since the frequencies range over numbers between zero and one, so will the degrees of confidence.

However, despite the attractiveness of such considerations, one must of course acknowledge that the Bayesian framework might be wrong. The crude ideas that it represents should not be controversial. However, it is quite possible that the Bayesian articulation of those ideas is not absolutely right; and that, in particular, the assumption of precise-valued, numerical degrees of belief is incorrect.

Even so, such a model could be an excellent idealization, sufficing perfectly well for the primary purposes of therapeutic Bayesianism, namely, to dispel confusion, solve problems, and thereby improve our understanding of the scientific method. For the paradoxes are caused by forgetting the *crude* facts (that there are gradations of belief, etc.) or by failing to recognize their significance. And so the solutions will involve noticing that those rough ideas have been overlooked and coming to appreciate how they bear on the problems. This sort of treatment will not depend essentially on any particular theoretical refinements. The function of the Bayesian framework is merely to cast the crude, uncontroversial ideas into a form where their impact on our problems can have maximal clarity and force.

What then is the import of studies that cast doubt on the existence of numerical degrees of conviction and which develop more complex and allegedly more realistic conceptions of belief? Let me stress that this work falls well outside the focus of therapeutic Bayesianism, for there is no reason to believe that such improvements will help to solve the standard problems in the philosophy of science. Perhaps these developments are important in psychology, statistics, semantics, or decision theory; perhaps they will become important to philosophy when we have progressed enough in our understanding of science so that the *details* of an inductive logic become items of reasonable concern. But at this juncture, confusion is rampant, the traditional problems are still very much with us, and it seems rather unlikely that the slight gains in accuracy to be derived from a more realistic theory of belief would be worth the price—in terms of loss of simplicity—that we would have to pay for it.

Our treatment of the 'raven paradox' is a case in point. The problem was solved by exposing a certain misconception (that a non-black, non-raven would be irrelevant to our hypothesis), and by explaining why we are so tempted by that misconception: namely, that in forgetting about degrees of belief, we lose sight of the distinction between very slight confirmation and no confirmation at all. The simple Bayesian model of belief provides a sufficiently perspicuous representation of the situation to enable us to put this in a clear way. Further accuracy regarding the nature of belief would distract us from the main point, ruin the argument, and not help us to understand the basis of the paradox.

Let me give another example. The conflict between realism and instrumentalism with respect to the acceptance of scientific theories is fueled by a shared tendency to think in terms of all-or-nothing belief. The instrumentalist argues, in light of previous scientific revolutions, that it is foolishly optimistic to expect that our current theories are true and will not eventually be refuted. Whereas the realist complains that it is a distortion of science to distinguish rigidly between credible observation reports and incredible theoretical claims. However, once we see that the issue is not 'To believe or not to believe?' but rather, 'To what degree

shall we believe?' then there is room for reconciliation. The crucial move is the elimination of the shared misconception. There is no reason to think that a fancy model of belief, even more accurate than the Bayesian idealization, would be any further help with the problem.

I do not mean to be suggesting that it is not worthwhile to investigate more sophisticated models of belief. On the contrary, I can readily imagine research programs—e.g., Bayesian psychology, or attempts to give a perfectly accurate description of scientific practice—in which this would be crucial. My point is that there is another enterprise—the one I am calling “therapeutic Bayesianism”—whose focus is on solving the traditional methodological puzzles and paradoxes, and for which the introduction of such complex models is likely to do more harm than good.

Suppose, then, that we do have numerical degrees of belief. Is there any way of justifying the Bayesian assumption that, to be rational, these degrees of belief must conform to the probability calculus? Although there are indeed various lines of reasoning which purport to establish this thesis, none is compelling. The best known of them is the ‘Dutch book’ argument⁸ and it goes roughly as follows. Defining a person’s degree of belief in a proposition as a function of the odds at which he is prepared to bet on its truth, it can be proved that if his degrees of belief do not satisfy the probability calculus then he will be prepared to accept a collection of bets which is guaranteed to lead to a loss. Therefore, since it would surely be irrational for him knowingly to put himself in such a no-win situation, it would be irrational to have a system of degrees of belief that violates the probability calculus. QED. However, the definition of ‘degree of belief’ that is employed in this argument presupposes that people maximize their expected utility. And there is a lot of room for skepticism about that assumption (and about the preference axioms to which it is equivalent). So the ‘Dutch book’ argument is far from airtight. Worse still, there is positive reason to think that its conclusion is false, for it requires logical omniscience. The probability of any logical truth is 1 and of any contradiction is 0. Yet it is surely quite rational to be less than perfectly confident in the truth of *some* logical truths—those that are especially hard to prove—and quite rational to give non-zero degrees of belief to contradictions that are hard to recognize as such.

The proper response to these difficulties is to repeat that the picture of rational degrees of belief obeying the probability calculus should be regarded as an *idealization* of the real normative facts. It is uncontroversial that one ought to be certain of elementary logical truths, and that one ought not be confident of the truth of obviously incompatible hypotheses. The probabilistic model of belief provides a sharp, perspicuous way of capturing these trivialities, and to the extent that it goes beyond them it need not be construed realistically.

A similar answer may be given to the third question concerning the definition of confirmation. In our discussion of the raven paradox we de-

defined “the degree by which E confirms H ” as “the ratio, $P(H/E)/P(H)$, for a rational person.” Evidently, this explication has at least some *prima facie* plausibility, and it certainly helps us to give a neat, compelling solution to the problem. Nonetheless, it is often argued that this particular explication is ‘wrong’—yielding counterintuitive consequences—and that there are better definitions of confirmation which should be used instead.⁹

However, these criticisms have little relevance to the project of therapeutic Bayesianism. No doubt our explication leads to some strange-sounding consequences. No doubt it is strictly speaking false that the ordinary meaning of “confirms” is given by our explication. No doubt there are definitions (perhaps involving non-probabilistic notions) that come closer to what we ordinarily mean. But the object of therapeutic Bayesianism is not to give a theory of science. We are not trying to find the most accurate analyses of our concepts, but rather to use explications that are at least roughly right, and which are conducive to simple, convincing dissolutions of philosophical problems. Since we assume that these problems are the product of confusion, it is desirable to look for ways of clarifying the issues, which have the proper blend of accuracy and simplicity. Of course it is possible to *oversimplify*. But one can conclude that this has happened only after finding that the admittedly idealized models do not in fact help to solve our problems.¹⁰

On the fourth question—Are there objective facts of confirmation?—it seems evident that judgments of credibility and confirmation do purport to capture objective normative facts. They do not state what any individual’s degrees of belief *actually* are, but rather they say something about what one’s degrees of belief *ought* to be, or how they *ought* to change given the circumstances. Thus we should acknowledge non-subjective facts regarding confirmation.

A natural way of capturing this idea, due to Carnap, is to suppose that an attribution of probability to a hypothesis reflects the belief in an objective, logical fact about the degree to which one statement—a summary of the available evidence—probabilifies [makes probable] another statement—the hypothesis in question. Such logical facts might be codified in a confirmation function, $c(p/q) = x$, which would specify explicitly the degree, x , to which q confirms p , and would specify implicitly the degree to which one should believe p if the total evidence is q . Carnap says, for example:

Probability-1 is the degree of confirmation of a hypothesis h with respect to an evidence statement e , e.g., an observation report. This is a logical, semantical concept. A sentence about this concept is based, not on observation of facts, but on logical analysis; if it is true, it is L-true (analytic). . . . Probability-2 [relative frequency] is obviously an objective concept. It is important to recognize that probability-1 is likewise objective.

Let h be the sentence 'there will be rain tomorrow' and j the sentence 'there will be rain and wind tomorrow'. Suppose someone makes the statement in deductive logic: ' h follows logically from j '. . . . The statement 'the probability-1 of h on the evidence e is $\frac{1}{5}$ ' has the same general character as the former statement. . . . Both statements express a purely logical relation between two statements. The difference between the two statements is merely this: while the first states a complete logical implication, the second states, so to speak, a partial logical implication; hence, while the first belongs to deductive logic, the second belongs to inductive logic.¹¹

Thus, Carnap held that certain facts about confirmation are analytic and *objective*, and thought of inductive probability as a *partial* version of the logical relation of entailment.¹²

On the fifth question—Does reason impose constraints on belief *over and above* the requirement of conformity with the probability calculus?—there are grounds for sympathy with Carnap's view that it does. For it is hard to see how the probabilistic constraint alone can account for our intuitions about the relative plausibility of competing hypotheses that equally well fit the current data. In particular, it is hard to see how it can solve the 'grue' problem.¹³

Suppose that such further constraints are indeed required. Still, to take up the sixth question, it is no trivial matter to say what they are. Carnap tried out various constraints and employed them to derive confirmation functions for certain extremely simple formal languages. Unfortunately, these functions have the counterintuitive property that laws of nature are never able to acquire more than a negligible probability. And this shows a deficiency in Carnap's constructions: either the languages are too simple, or the constraints are wrong. However, one can certainly not conclude that anyone who endorses a Carnapian conception of logical probability *must* hold that general laws never attain a non-negligible probability. This is a non sequitur, arising from a failure to distinguish between the general conception of logical probability and the admittedly inadequate prototypes with which Carnap experimented.¹⁴

For the treatment of various problems it is helpful to suppose that inductive reasoning is represented by a specific (but unspecified) real-valued Carnapian confirmation function, c , allowing general laws to achieve a nonnegligible credibility. In light of our responses to questions (1), (2), and (3), we see that it can be no objection to this procedure that our inductive practice is not in fact precisely described by a single c -function. For, once again, the intention is not to get at the exact truth, but merely to employ a useful idealization. Nor—as we have just said—is it fair to complain that *some* c -functions—those Carnap toyed with—always give zero probability to general laws. For we can suppose that 'the right c -function' is one of those that do *not* have that counterintuitive feature.

III | Misplaced Scientism

Much criticism of therapeutic Bayesianism arises from a conflation of philosophy and science. More exactly, it derives from a failure to recognize the legitimacy (even the existence) of non-scientific philosophical projects—those prompted, not by a desire to expose the whole truth regarding some domain, but by an interest in the resolution of paradoxes. Let me elaborate.

What one might call '*theory-oriented* philosophy of science' aims for a systematic account of the scientific method. The criteria of success are just those that pertain to theory construction *within* particular sciences: namely, empirical adequacy, scope, depth, simplicity, internal consistency, and coherence with the rest of our knowledge. More specifically, a perfect theory of the scientific method would be expected to conform with specific intuitions about the way that good science is done, to cover all aspects of methodology in detail, to expose fundamental principles enabling the complex, superficial aspects of scientific practice to be unified and explained, and to respect results in psychology and sociology. Thus it seems appropriate to regard theory-oriented philosophy of science as itself a department of science—a branch of naturalized epistemology. This characterization neither ignores nor denies that scientific methods are normative. A description of science will contain a codification of the basic norms which are implicit in the evaluation of theories. Moreover, it is quite possible that an identification of basic normative principles will result in the exposure of cases in which science is being done badly. One might thereby effect an improvement in the conduct of some science. Indeed, this may well be a motive for engaging in theory-oriented philosophy of science.

In contrast, the approach portrayed in this paper, '*problem-oriented* philosophy of science', has very different goals, methods, and adequacy conditions. It aims at the resolution of deep puzzles and paradoxes that arise from reflection upon science. It includes in its domain, for example, the problem of induction, the paradox[es] of confirmation, the question of total evidence, and the issue of prediction versus accommodation. The problems here are not simply to fill various undesirable gaps in our knowledge about science. Characteristically, they are conceptual tensions, contradictions, absurd conclusions—that is to say, symptoms of confusion. We have somehow gone astray, and the task is to understand how this has happened and to get a clear view of the issue so that our misguided ways of thinking will be exposed and no longer seem so attractive.

These two approaches to the philosophy of science do not compete with one another. They are distinct projects with distinct objectives—not *wholly* unrelated to one another, but by no means simply parts of the same enterprise. Thus, it is not the case that the sort of full understanding

provided by a successful theory-oriented philosophy of science would automatically solve the puzzles that form the domain of problem-oriented philosophy of science. For the resolution of a paradox requires a great deal more than just locating the wrong move in a fallacious argument. It is crucial to a proper resolution that one comes to see why that fallacy was natural. And it is important that one obtains a new perspective on the issue—a point of view from which the old and troublesome habits of thought no longer seem plausible. These elements of the solution do not simply fall out of a complete theory of science. (Similarly, the explanation of conjuring tricks does not follow from physics.) Thus, theory-oriented philosophy of science is not simply a more thorough, systematic, and ambitious project than problem-oriented philosophy of science.

Neither is it necessary, in order to solve problems, that one be in possession of an adequate theory of science. For confusions can be identified, understood, and removed without a theory of any particular depth or generality. Granted, assumptions about methodology will often be involved in the diagnosis and treatment of a problem, and if these were *wildly* false then it is unlikely that the discussion would be helpful. However, there is no reason why such assumptions should be *true* as long as their replacement with the truth would not undermine the solution that is based on them. Indeed it is quite possible that the perfect theory of science would be a very bad tool for solving problems. For the truth may be so complicated that it cannot provide the sort of simple and relevant perspective that is needed.

If the practice of conceptual troubleshooting is confused, as it often is, with the scientific search for a theory of science, then therapeutic Bayesianism will be wrongly subjected to all of the methodological requirements that are properly applied only in science. Let me describe some of the bad effects of this confusion.

One consequence, discussed above, of not seeing the distinctive aim of therapeutic Bayesianism is a tendency to misjudge the function of various helpful idealizations. Thus one commonly finds objections to the use of precise-valued degrees of belief, to the assumption that these should conform to the probability calculus, to the adoption of a particularly simple explication of confirmation, and to the idea that our inductive practice may be represented by a single Carnapian confirmation function. Doubtless some of these assumptions are, strictly speaking, false. (Just as it is false that a gas is made of point masses.) And in a different kind of study—one aimed at truth—it would be very important to discuss more realistic models. However, for the purposes of therapeutic Bayesianism it is important to use the simplest roughly accurate models of degrees of belief and of confirmation that will help to clarify the issues, and it is sufficient to proceed on the basis of their intuitive plausibility and to justify these models in retrospect in terms of their utility.

Secondly, a scientific understanding of confirmation aims for the

truth, the *whole* truth, and nothing but the truth. Consequently, those wedded to this conception of the philosophy of science will find fault with studies that do not discuss every significant aspect of the phenomenon of confirmation. Consider, for example, *prior probability assignment*, the procedures for deciding, before data have been gathered, the various 'intrinsic' plausibilities of hypotheses; *belief-kinematics*, the way that systems of belief change over time in the light of new discoveries; or *direct inference*, the impact on our degrees of belief of a knowledge of empirical probabilities. These are fascinating topics, and a good theory of science must deal with them. But there is no reason why a paradox-oriented Bayesian program should incorporate a complete, systematic account of all such elements of methodology.¹⁵

In the third place, scientism in philosophy engenders a 'hyperformalist' fixation on symbolic technique—an overvaluation of logico-mathematical machinery. Among the symptoms of this hyperformalist state are: (a) a blindness to the possibility of philosophical problems distinct from the scientific and mathematical issues that arise in statistics, decision theory, sociology of science, etc., further questions being dismissed as 'merely verbal';¹⁶ (b) a dissatisfaction with informal discussions and conclusions; (c) an exaggerated concern with formal rigor for its own sake; and (d) an obsession with the elimination of any potential ambiguity or vagueness, leading to the feeling that the English language is too confusing and vague a medium for intellectual progress, and that it should, wherever possible, be replaced with mathematics or logic.

Thus, even if an approach employs formal techniques, as therapeutic Bayesianism clearly does, it may still be subjected to hyperformalist criticism. I think this is an unhealthy point of view—in philosophy generally, and particularly in the philosophy of science, where it is especially common. No doubt there are occasions when clarity is gained and confusion allayed with the help of formal apparatus. This, I believe, is one of the morals of Bayesianism's success. However, one can withdraw too quickly into the secure, regulated territory of a formal system. It is certainly a tempting relief from the frustrating vagaries of philosophy to be able to obtain definite, proven results and get clear answers to clear questions. But, unless we are very careful, these answers and results might have little to do with the problems that have traditionally motivated philosophy of science. Our methodological puzzles arise when we reflect informally about scientific practice; and they can be solved only with an appreciation of the misconceptions and confusions to which we are prone and an understanding of the ways in which they are fostered by the rich conceptual resources put at our disposal by natural language. It seems to me that only when that sort of understanding is eventually attained will we know what we are looking for in a fully fledged inductive logic; and then, perhaps, be in a better position to devise one. But this level of understanding will not be achieved by trying to express as many questions as possible within

a formal system, proving some theorems, and dismissing the residue as intractable and uninteresting. At its worst, such scientistic hyperformalism betrays a lack of concern for truly philosophical problems. If “merely verbal” issues are any that do not make a scientific difference, and if only scientific problems are worth worrying about, then philosophy is truly an endangered enterprise.

I hope to have clarified what I believe is a valuable approach to the philosophy of science, and to have shown that many of the complaints about it derive from scientistic hyperformalism and are therefore misconceived. The goal is not a theory of science but the unravelling of puzzles surrounding our ideas about surprising data, prediction versus accommodation, ad hoc postulates, statistical hypotheses, our thirst for new data, the tenability of realism, and other aspects of methodology. And given some of the successes of therapeutic Bayesianism there is reason to have a fair amount of confidence in its basic principles.

Thus, the notion of rational degrees of belief conforming to the probability calculus has an important role in the philosophy of science. It would no doubt be easier to think in terms of all-or-nothing belief, but that oversimplification is part of what engendered our methodological puzzles in the first place. On the other hand, there are more complex and realistic conceptions of belief, but the cause of clarity is not served by using them. Therapeutic Bayesianism appears to offer the ideal compromise between accuracy and simplicity, enabling us to represent the issues starkly without neglecting the essential ingredients or clouding them with unnecessary details.¹⁷

■ | Notes

1. The axioms of elementary probability theory are as follows: (1) probabilities are less than or equal to one; (2) the probability of a necessary truth is equal to one; (3) if two statements are jointly impossible, then the probability that at least one of them is true is equal to the sum of their individual probabilities; and (4) the conditional probability of p given q equals the probability of the conjunction of p and q divided by the probability of q .
2. This project is attempted in my *Probability and Evidence* (Cambridge, 1982), henceforth abbreviated as *P&E*. The metaphilosophical outlook is inspired by Wittgenstein's *Philosophical Investigations*, paragraphs 88–133.
3. Bayesian programs of various kinds have been developed in the work of Rudolf Carnap, David Christensen, R. T. Cox, Bruno de Finetti, Ron Giere, I. J. Good, John Earman, Ellery Eells, Hartry Field, Allan Franklin, Ian Hacking, Mary Hesse, Jaakko Hintikka, Colin Howson, E. T. Jaynes, Richard Jeffrey, Harold Jeffreys, Mark Kaplan, J. M. Keynes, Henry Kyburg, Isaac Levi, Patrick Maher, Roger Rosenkrantz, Wesley Salmon, L. J. Savage, Teddy Seidenfeld, Abner Shimony, Brian Skyrms, Patrick Suppes, Peter Urbach, Bas van Fraassen and others. Much of this

work (especially the studies by Good, Hesse, Howson & Urbach, and Earman) contains contributions to therapeutic Bayesianism. However, I cannot attribute to these philosophers the project that I have in mind by that label, since their work is oriented towards the discovery of a 'theory of science', and thus reflects a metaphilosophical point of view that is quite distinct from that of the program which I am calling "therapeutic Bayesianism."

4. Stephen Spielman's objection is based on the mistake described here: the identification of the probabilities with objective proportions. (See his review of *P&E*, *Journal of Philosophy* 81 [March 1984], 168–73. Page references for Spielman are to this work).

To keep things relatively simple I have assumed that there are just two observations in question: namely, the discovery regarding a randomly selected raven that it is black and the discovery regarding a randomly selected non-black thing that it is not a raven. If we consider instead the discovery that a known black thing is a raven, or various other ways of seeing black ravens and non-black non-ravens, then the existence of confirmation depends on the presence of special additional background assumptions (e.g., that ravens are quite likely all to have the same color). Nonetheless a similar contrast between the degrees of confirmation provided by black ravens and non-black non-ravens may be established. In *P&E* I suggest that these other ways of seeing black ravens would provide *no* confirmation of the hypothesis. This is misleading. Sometimes our background theories include a belief in the projectibility of the generalization in question, and in that case all the ways of observing an instance of it will normally provide confirmation.

My treatment of the paradox is, in a couple of respects, different from Patrick Suppes's analysis ("A Bayesian Approach to the Paradoxes of Confirmation," in *Aspects of Inductive Logic*, edited by J. Hintikka and P. Suppes, [Amsterdam 1966]). In the first place, he does not distinguish between the discovery that a randomly selected object is a black raven and the discovery that a randomly selected raven is black; whereas it is a significant feature of my account that in certain circumstances only the latter datum would confirm the hypothesis. And secondly, he does not obtain his results from the basic principles of Bayesianism—the thesis that degrees of belief should conform to the probability calculus; rather, he starts with the assumption that surprising observations have greater confirmation power; and this, though correct, is much better derived than simply presupposed.

5. C. Howson and P. Urbach, *Scientific Reasoning: The Bayesian Approach* (La Salle, Ill., 1989)

6. An argument for associating simplicity with high prior probability is given in *P&E*, 70–71.

7. Teddy Seidenfeld maintains that this account goes in the "wrong direction." But he gives no grounds for that claim other than to note the above-mentioned deficiencies in our understanding of simplicity—our inability to solve either the descriptive or the normative problems surrounding it. And it seems to me that his observation is irrelevant in the absence of any reason to believe either (a) that we can get a satisfactory explication of simplicity in terms of evidential diversity, or (b) that the Bayesian account would not withstand a better grasp of simplicity. (See his review of *P&E*, *Philosophical Review*, 93 [July 1984] 474–83.)

8. For a good assessment of this argument and various others see John Earman's *Bayes or Bust?* (Cambridge, Mass., 1992). Bruno de Finetti perhaps deserves the credit for first having argued that degrees of belief ought to be 'coherent', i.e., conform to the probability calculus—though they need not be coherent if the believer is irrational ("Foresight: Its Logical Laws, Its Subjective Sources," translated in *Studies in Subjective Probability*, edited by H. E. Kyburg, Jr., and H. E. Smokler [New York, 1964]). I hesitate to credit Frank Ramsey's earlier paper (in *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*, edited by D. H. Mellor [Atlantic Highlands, N.J., 1977]) with this result, since he defines "degrees of belief" in such a way that they must conform to the probability calculus. On Ramsey's account there is no room for the existence of someone who has degrees of belief that are not coherent.

9. For example, I. J. Good (in the *British Journal for the Philosophy of Science* 19 [1968]: 123–43) advocates:

$$\text{Weight of evidence concerning } H \text{ provided by } E = \log \frac{P(E/H)}{P(E/-H)}$$

And Seidenfeld (op. cit.), noting that on our account E might confirm both H_1 and H_2 yet disconfirm the conjunction ($H_1 \& H_2$), suggests that confirmation cannot be defined in terms of probability alone.

10. A further complaint is that our definition of confirmation seems to go badly wrong when we apply it to measure the evidential value of *already known* data. For in that case $P(E) = 1$, therefore $P(H/E) = P(H)$. This problem for Bayesians was first posed by Clark Glymour (see his *Theory and Evidence* [Princeton, 1980]). It has been forcefully reiterated by James Woodward (in his review of *P & E*, *Erkenntnis* 23 (1985): 213–19) and treated thoroughly by John Earman (in *Bayes or Bust?*). In order to deal with it we should remember that the idea of the definition is to compare the credibility of a hypothesis, H , given the knowledge that E is true, with its credibility in the absence of such knowledge. Thus we should take the prior probability to be that which H would have had if the truth of E had not been discovered. Then, in order to assess E 's confirmation power, we should consider what the absolute subjective probability of E would have been in that counterfactual situation, and also what the conditional probability of E given H would have been. Then we can employ Bayes's Theorem to calculate the factor by which the prior probability of H would have been increased. Doubtless, there is substantial indeterminacy in the assessment of these counterfactual probabilities. But this is no objection, since we generally have no reason to expect the magnitude of E 's confirmation power to be an especially determinate matter.

11. Rudolf Carnap, *Logical Foundations of Probability* (Chicago, 1962), 19, 31.

12. According to Spielman, this construal of Carnap is a "distorted caricature" (170), for "any careful reading of LFP [Logical Foundations of Probability] would show that Carnap never talks about 'objective relations of probabilification' or 'objective' relations of partial entailment" (170). Here I am at a loss to explain how Spielman could have arrived at his interpretation, and I can only refer the reader back to Carnap's work.

13. For further discussion of this point see *P & E*, 32–36 and 74–81 and Earman's *Bayes or Bust?*, chapter 6.

14. Spielman (171) falls into this error, complaining that one cannot endorse logical probability and yet still assume that laws can have a non-negligible credibility.

15. Thus Woodward writes: "The principal defect of *Probability and Evidence* is its unsystematic character. Horwich does not give us a fully worked out general theory of confirmation but rather a series of essays which offer solutions to various particular puzzles, where the interconnections among these solutions are by no means always clear" (214).

16. This is starkly revealed in Seidenfeld's dismissal of therapeutic Bayesianism on the grounds that it is no substitute for a combination of excellent, but highly technical, foundational studies in decision theory and statistical inference by Jeffrey, Fishburn, and Lindley—works that hardly touch upon the traditional philosophical puzzles that form the domain of therapeutic Bayesianism. In a similar vein, Spielman is bothered by the "fail(ure) to see that the only difference between an 'objectivist' account [of the 'grue' problem] and a personalist account would be verbal: an objectivist would say that we *ought* to assign a much higher probability to H_1 than to H_2 , and a subjectivist says that this is what intelligent informed people in fact do" (170). Spielman thinks the issue between them is 'merely verbal'.

17. I have greatly benefited from James Woodward's thorough and perceptive criticism. I would also like to thank Ned Block, Susan Brison, Josh Cohen, Marcus Giaquinto, Mark Kaplan, and Judith Thomson for helping me to improve earlier drafts of this paper.

5.1 Bayes for Beginners 627

The Relevance Criterion of Confirmation 627

Bayes's Theorem and the Axioms of Probability Theory 628

Bayes's Theorem and Scientific Reasoning 632

Probabilities and Degrees of Belief 634

5.2 Salmon on Kuhn and Bayes 638

The Inadequacies of Hypothetico-Deductivism 638

Prior Probabilities 640

The Expectedness of Evidence 642

Likelihoods and the Catch-All Hypothesis 643

Salmon's Bayesian Algorithm for Theory Preference 644

5.3 Why Glymour Is Not a Bayesian 646

Bayesian A Priori Arguments 647

Methodological Truisms 650

Simplicity 651

The Relevance of Evidence to Theory 654

The Problem of Old Evidence 656

5.4 Horwich's Defense of Therapeutic Bayesianism 659

The Raven Paradox 659

Diversity of Evidence 663

Probabilistic Foundations 665

Misplaced Scientism 665

5.5 Summary 667

5 | COMMENTARY

5.1 | Bayes for Beginners

Although the Bayesian approach to confirmation is not inconsistent with the hypothetico-deductive (H-D) approach (discussed in chapter 4), it differs from it in several fundamental respects. The aim of this first section is to explain and illustrate those differences as a prelude to the analysis of the readings in chapter 5 later in this commentary.

THE RELEVANCE CRITERION OF CONFIRMATION

A distinctive feature of the Bayesian approach is its reliance on the mathematical theory of probability. Unlike the H-D approach, which treats confirmation as a qualitative notion that might be made quantitative later on, Bayesians assume that confirmation is quantitative from the outset. Even such qualitative notions as evidence confirming a hypothesis and evidence confirming one hypothesis more strongly than it does another are analyzed in terms of probabilities with numerical values that lie between 0 and 1. Bayesians contend that this essentially quantitative approach to confirmation in terms of probability theory can solve the puzzles and paradoxes afflicting purely qualitative theories of confirmation, such as Hempel's satisfaction criterion.

The most common way that Bayesians connect confirmation with probability is by adopting the relevance criterion of confirmation, according to which a piece of evidence, E , confirms a hypothesis, H , if and only if E raises the probability of H :

Relevance Criterion of Confirmation	E confirms H if and only if $P(H/E) > P(H)$; E disconfirms H if and only if $P(H/E) < P(H)$.
--	---

For convenience, we shall often refer to $P(H)$ as the *prior probability* of H , and $P(H/E)$ as the *posterior probability* of H . $P(H/E)$ is a conditional probability and should be read as the *probability of H given E* .

It should be noted that, although it is defined in terms of quantitative probabilities, the notion of confirmation (often called *incremental confirmation*) defined by the relevance criterion is qualitative. The relevance criterion does not specify how degrees of confirmation should be measured; it merely gives a necessary and sufficient condition for that confirmation. Indeed, there is an ongoing dispute in the literature about whether numerical degrees of confirmation should be a function of the ratio of $P(H/E)$ to $P(H)$ or a function of the difference between them.¹ Regardless

of where they stand on this issue, all Bayesians agree that the more E raises the probability of H , the more E confirms H .

The relevance criterion of confirmation differs significantly from the absolute criterion of confirmation, according to which E confirms H if and only if $P(H/E)$ exceeds some suitably high threshold value, say, 0.9. From the point of view of those who endorse the relevance criterion, the absolute criterion confuses confirmation with acceptance. High probability may be an appropriate condition for accepting a hypothesis, but it is not necessary for confirmation. Thus, adherents to the relevance criterion would consider H confirmed by E even though E raised the probability of H only a little, from, say, 0.2 to 0.4, and the posterior probability of H , $P(H/E)$, remained less than 0.5.

BAYES'S THEOREM AND THE AXIOMS OF PROBABILITY THEORY

Bayes's theorem (also called Bayes's rule, law, or equation) lies at the heart of the Bayesian approach to confirmation and gives that approach its name. In this section we shall be concerned solely with Bayes's theorem as a formal result in probability theory. As such, Bayes's theorem, like any mathematical theorem, is entirely uncontroversial. What is distinctively Bayesian about the Bayesian approach to confirmation is not merely its use of Bayes's theorem but its interpretation of the probabilities occurring in the theorem. The Bayesian interpretation of probabilities as subjective degrees of belief will be discussed later, in the section "Probabilities and Degrees of Belief."

Bayes's theorem is a deductive consequence of the three basic axioms of probability theory. Everything else in probability theory can also be deduced from these axioms, supplemented with definitions of notions such as conditional probability. Here are the axioms in their unconditional form.

- Axiom 1 Every probability is a real number between 0 and 1:
 $0 \leq P(A) \leq 1$.
- Axiom 2 If A is a necessary truth, then $P(A) = 1$.
- Axiom 3 If A and B are mutually exclusive (that is, if it is impossible for both A and B to be true), then $P(A \vee B) = P(A) + P(B)$.
 This theorem is often referred to as the *special addition rule*.

Strictly speaking, the A , B , C , and so on that probability ranges over are propositions, but we shall, when convenient, talk about the probability of events, theories, classes of theories, and evidence.

Even though the set of axioms is small, several important rules that we shall use later on can be deduced from them.

Negation Rule	$P(\sim A) = 1 - P(A)$.
Implication Rule	If A logically entails B , then $P(B) \geq P(A)$.
Equivalence Rule	If A and B are logically equivalent, then $P(A) = P(B)$.
General Addition Rule	$P(A \vee B) = P(A) + P(B) - P(A \& B)$.

The general addition rule is especially useful, since it applies regardless of whether A and B are exclusive. Obviously, when A and B are mutually exclusive, $P(A \& B)$ is 0 and the general addition rule reduces to the special addition rule (axiom 3).

One way to make the general addition rule intuitively obvious is to represent propositions by circles and to let the probability of each proposition equal the area of its circle.² When A and B are mutually exclusive, the A -circle and the B -circle do not overlap, and the probability of $(A \vee B)$ is simply the sum of $P(A)$ — the area of the A -circle — and $P(B)$ — the area of the B -circle (figure 1). When A and B are not exclusive, the circles overlap (figure 2).

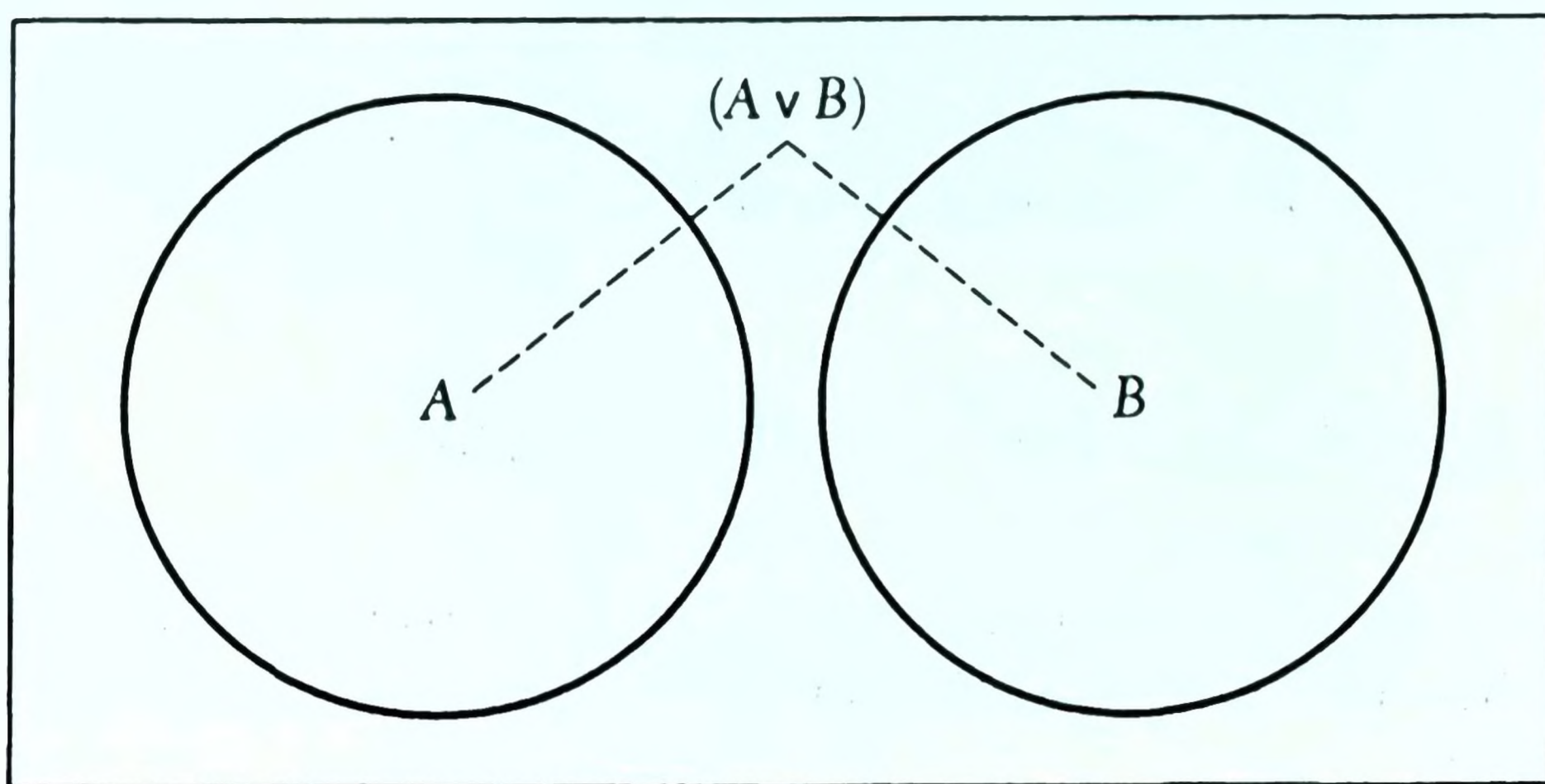


Figure 1

Thus, to calculate $P(A \vee B)$ for figure 2 we add the areas of the two circles as before, but then we have to subtract $P(A \& B)$, which is the area of the overlap, to get the correct answer.³

To derive Bayes's theorem from the probability axioms, we need a definition of $P(A/B)$, the conditional probability of A given B .⁴ It is:

Definition of Conditional Probability

$$P(A/B) = \frac{P(A \& B)}{P(B)}, \text{ where } P(B) > 0.$$

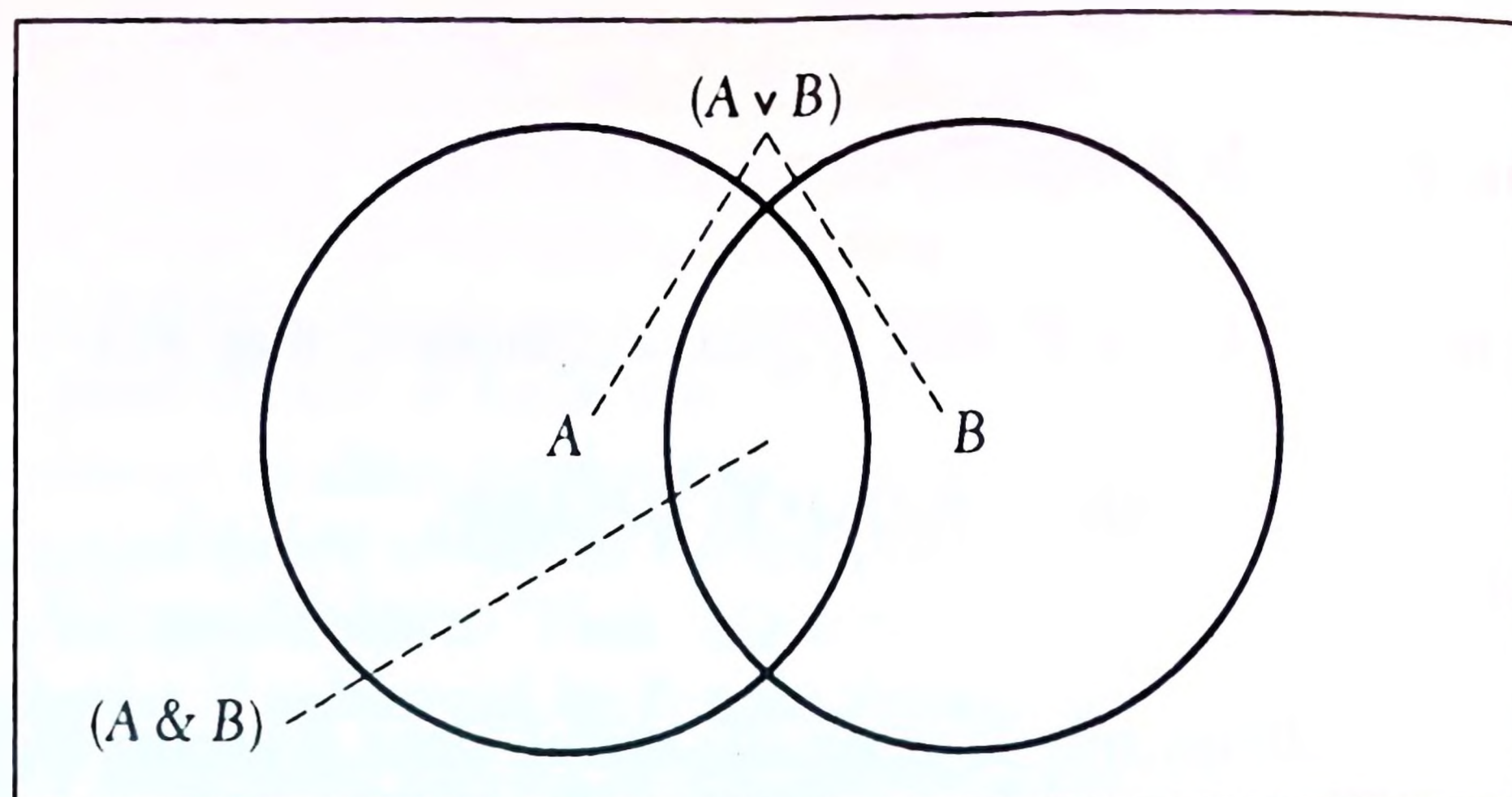


Figure 2

The rationale for adopting this definition of conditional probability can be appreciated by considering figure 2. Suppose that you are told that a dart has been thrown, randomly, at the figure and has landed somewhere inside the B -circle. Given that the dart is inside the B -circle, what is $P(A/B)$, the probability that the dart is also inside the A -circle? The answer is simple: it is the area common to both circles divided by the area of the B -circle. In other words, given that probabilities are proportional to areas, $P(A/B)$ is equal to $P(A\&B)$ divided by $P(B)$.

It is an immediate consequence of our definition of conditional probability that a general multiplication rule holds for the probability of any conjunction (or for the probability of the joint occurrence of any two events).

General Multiplication Rule $P(A\&B) = P(A/B) \times P(B)$.

When $P(A/B) = P(A)$, A and B are said to be statistically independent of one another and the general multiplication rule simplifies to the special multiplication rule.

Special Multiplication Rule When A and B are independent,
 $P(A\&B) = P(A) \times P(B)$.

It is only a short step from the general multiplication rule to Bayes's theorem. First, we note that, since $(A\&B)$ is logically equivalent to $(B\&A)$, it follows from the equivalence rule that:

$$P(A\&B) = P(B\&A).$$

Substituting, using the general multiplication rule, gives:

$$P(A/B) \times P(B) = P(B/A) \times P(A).$$

Rearranging the terms gives the simplest form of Bayes's theorem:

$$\text{Bayes's Theorem} \quad P(B/A) = \frac{P(A/B) \times P(B)}{P(A)}, \text{ where } P(A) > 0.$$

Before trying to apply Bayes's theorem to scientific reasoning and the confirmation of theories by evidence, let us consider a simple example that illustrates its essential features. Imagine that a make of wheelchair, the Samson, is manufactured in just two plants in the United States. One factory is in Boston, the other in Chicago. The Boston plant makes four-fifths of all Samsons; the Chicago plant makes the rest. Of the Samsons manufactured in Boston, one-sixth have a special lightweight aluminum frame, whereas, three-quarters of the Samsons made in Chicago are of this type. You purchase a Samson wheelchair at an auction and discover that it has a lightweight aluminum frame. What is the probability that it was made in Boston?

Let A stand for having a lightweight aluminum frame, let B stand for being made in Boston, and let C stand for being made in Chicago. We want to calculate $P(B/A)$ using Bayes's theorem. The information we are given is that $P(B) = 4/5$, $P(C) = 1/5$, $P(A/B) = 1/6$, and $P(A/C) = 3/4$. Bayes's theorem tells us that

$$P(B/A) = \frac{P(A/B) \times P(B)}{P(A)}.$$

Thus, the numerator equals $1/6 \times 4/5 = 2/15$. But what about the denominator, $P(A)$? Samsons are made in only two places—Boston and Chicago—so the alternatives are $(A \& B)$ and $(A \& C)$, which are mutually exclusive and exhaustive. Thus, by axiom 3,

$$P(A) = P[(A \& B) \vee (A \& C)] = P(A \& B) + P(A \& C),$$

and then, using the general multiplication rule, we get:

$$\begin{aligned} P(A) &= P(A/B) \times P(B) + P(A/C) \times P(C) \\ &= (1/6) \times (4/5) + (3/4) \times (1/5) = 17/60. \end{aligned}$$

Thus:

$$P(B/A) = (2/15) \times (60/17) = 8/17.$$

When using Bayes's theorem, we need not calculate the denominator, $P(A)$, from first principles every time. Instead, we can write down the answer immediately using the total probability rule:

Total Probability Rule $P(A) = P(A/B) \times P(B) + P(A/\sim B) \times P(\sim B)$.

In the wheelchair example, there were just two exclusive alternatives: either the chair came from Boston (B) or from Chicago (C). So, in this example, we can calculate $P(A)$ by substituting C for $\sim B$ in the total probability rule:

$$P(A) = P(A/B) \times P(B) + P(A/C) \times P(C).$$

But it is easy to imagine a more complicated example in which aluminum chairs were also made in Detroit (D) and Evanston (E). In this case, the right-hand side of the equation for $P(A)$ will include two extra factors:

$$P(A/B) \times P(B) + P(A/C) \times P(C) + P(A/D) \times P(D) + P(A/E) \times P(E).$$

More generally, when B_1, B_2, \dots, B_n are n mutually exclusive and exhaustive hypotheses, we can express $P(A)$ as the sum of products as follows.

$$P(A) = P(A/B_1) \times P(B_1) + P(A/B_2) \times P(B_2) + \dots + P(A/B_n) \times P(B_n).$$

This can be written more succinctly as:

$$P(A) = \sum_{i=1}^{i=n} P(A/B_i) \times P(B_i).$$

BAYES'S THEOREM AND SCIENTIFIC REASONING

The wheelchair example used above to illustrate Bayes's theorem can be regarded as a simple analogue of scientific reasoning. Let the theory, T , be that our Samson wheelchair was made in Boston. Initially, before we discovered that the chair is made of aluminum, the probability that it came from Boston was $4/5$, since 80 percent of all Samson wheelchairs are made there. In other words, in our example the prior probability of theory T , $P(T)$, is $4/5$. Once we acquire evidence, E , that the chair is made of aluminum, we calculate that the posterior probability of T given evidence E , $P(T/E)$, is $8/17$. So the probability of T has dropped, and E disconfirms T . (In fact, in this example, it is now slightly more probable than not that the chair came from Chicago rather than Boston.)

In light of this example, we can summarize the application of Bayes's theorem to scientific theories as follows:

Bayes's Equation
(Version 1)

$$P(T/E) = \frac{P(E/T) \times P(T)}{P(E)}$$

$P(T)$ is the prior probability of T , $P(E)$ is the probability of the evidence E (what Salmon calls the *expectedness of the evidence*), and $P(T/E)$, the probability of T conditional on E , is the posterior probability of T . The meaning of these terms is pretty straightforward and obvious, but the terminology for $P(E/T)$ can be misleading. $P(E/T)$, the probability of E conditional on T , is usually referred to as the *likelihood of T* (or, as some authors prefer, the *likelihood of T on E*). To repeat, the likelihood of T is $P(E/T)$, not $P(T)$, and similarly, the likelihood of T on E is $P(E/T)$, not $P(T/E)$. To avoid possible confusion, we will avoid using the term *likelihood* as a synonym for *probability* and use it solely to refer to $P(E/T)$ and similar expressions.⁵

When given some new evidence E for theory T , we revise our assessment of the theory's probability by using Bayes's equation to calculate $P(T/E)$ as a function of $P(T)$, $P(E/T)$, and $P(E)$. We then discard our old prior probability, $P(T)$, and replace it with $P(T/E)$. In this way, the posterior probability of T becomes our new prior probability:

$$P_{\text{new}}(T) = P(T/E)$$

Some authors, such as Salmon, explicitly include background knowledge, B , in Bayes's equation, thus making all the probabilities involved conditional. In the wheelchair example, B would include the information that the chairs are made either in Boston or Chicago and nowhere else. In real-life scientific reasoning, B would include the information about the world and other theories that scientists accept as true or highly probable. Including background knowledge, the equation becomes:

Bayes's Equation
(Version 2)

$$P(T/E\&B) = \frac{P(E/T\&B) \times P(T/B)}{P(E/B)}$$

We can also use the total probability rule to expand the denominator as follows.

$$P(T/E\&B) = \frac{P(E/T\&B) \times P(T/B)}{P(E/T\&B) \times P(T/B) + P(E/\sim T\&B) \times P(\sim T/B)}$$

Since it is cumbersome to write out these equations when every probability is explicitly conditionalized upon the background information, we shall omit the reference to B whenever it is convenient. Thus, Bayes's equation can be expressed more simply as:

Bayes's Equation
(Version 3)

$$P(T/E) = \frac{P(E/T) \times P(T)}{P(E/T) \times P(T) + P(E/\sim T) \times P(\sim T)}$$

And in its most general form, when T_1, T_2, \dots, T_n are mutually exclusive and exhaustive, we have:

Bayes's Equation
(Version 4)

$$P(T_k/E) = \frac{P(E/T_k) \times P(T_k)}{\sum_{i=1}^n P(E/T_i) \times P(T_i)}$$

PROBABILITIES AND DEGREES OF BELIEF

In order to apply Bayes's equation (in any of its four versions) to scientific reasoning, we have to decide what the probabilities that appear in it mean and how they should be measured. The pure Bayesian line is that, subject to some important conditions, all probabilities are subjective degrees of belief and that rationality requires us to revise our beliefs using Bayes's equation. The subjective (or so-called personalist) interpretation of probability and the alleged connection between rationality and Bayes's equation are controversial issues that are debated in the readings in this chapter. The purpose of this section is to explain and clarify the main issues involved in that debate.

The right-hand side of Bayes's equation (version 1) contains three sorts of probabilities: $P(T)$, $P(E)$, and $P(E/T)$ —the prior probability of theory T , the expectedness of evidence E , and the likelihood of T . In many respects, likelihoods are the least problematic, since, even if $P(E/T)$ is a subjective degree of belief, it seems rational to base that probability on the objective relation between T and E . For example, if T is a deterministic theory that deductively entails E , then everyone agrees that the correct value to be assigned to $P(E/T)$ is 1. Similarly, if T is a statistical theory, as in our wheelchair example, then T (together with background information B) will specify the probability that E is true on the condition that T is true.

Assigning values to $P(E)$ and $P(T)$ is more difficult. Later, in the section on Glymour, we discuss the so-called problem of old evidence, that is, the problem of assigning to $P(E)$ some value other than 1 when E is already known to be true. Of course, one might try to use the second or third versions of Bayes's equation to calculate $P(E)$, but as Salmon points out in his article, to do so we would need to calculate $P(E/\sim T)$, and since $\sim T$ is simply the negation of T , not a specific theory, we cannot infer from $\sim T$ the value of $P(E/\sim T)$. The fourth version of Bayes's equation avoids the indeterminate character of $\sim T$ but only at the price of requiring a complete list of all the possible theories that predict E (or that assign to E some definite probability). In practice, only a small handful

of rivals to T (possibly none) will be candidates for serious consideration. Scientists simply do not know what all the logically possible rival theories are. For further discussion of this problem, see the articles by Salmon and Glymour and the sections on their articles later in this commentary.

Even if values for $P(E/T_i)$ were available for every single theory T_i that predicts E (either with certainty or some definite probability), the problem of determining the prior probabilities, $P(T_i)$, of each of these theories would remain. At this point subjectivist Bayesians say that the prior probability of a theory, T , is simply the actual degree of belief that a person has in T . Strictly speaking, on the subjectivist interpretation of probability, there is no such thing as *the* prior probability of T , since degrees of belief are relative to persons and it is perfectly possible that different people believe the same theory to different degrees, even given the same background information. According to subjectivist Bayesians, it is an illusion to think that there is one objectively correct answer to the question, "What is the prior probability of T ?"

It seems incredible, on the face of it, that anything useful could be said about scientific inference and confirmation, starting from a basis as subjective as each person's degree of belief. In the remainder of this section, we take a brief look at some of the most important objections to the subjective interpretation of probability and the Bayesian replies to these objections. Many of these issues are explored more fully in Glymour's article and discussed later in this commentary.

Bayesians begin with degrees of belief. What are they, and how can they be measured so as to yield numerical probability values? The pioneers of Bayesian theory (Frank P. Ramsey, Bruno De Finetti, Leonard J. Savage) interpreted degrees of belief in terms of people's behavior. Ramsey proposed, for example, that a person's degree of belief in any proposition be measured by the least odds at which he would be willing to gamble on the proposition being true. In this way, we connect degrees of belief with something we can observe and measure, namely, betting behavior. But even when they are measured in this way, why should we think that the degrees of conviction that a person happens to have in various propositions qualify as probabilities by satisfying the probability axioms? Surely, people violate the axioms in many cases. For example, there might be a proposition Q , such that a person's degree of conviction in Q and his degree of conviction in $\sim Q$ do not add up exactly to 1.

The Bayesian response to this objection is the Dutch book argument. Professional gamblers say that a Dutch book has been made against someone if that person accepts a series of bets such that, no matter what the outcome, the person is guaranteed to lose money. No rational person would knowingly gamble in this way. The Dutch book theorem proves that a necessary and sufficient condition for avoiding a book being made against you is that your degrees of belief satisfy the axioms of probability theory. When this condition is satisfied, your degrees of belief are said to

be coherent. Thus, when subjective Bayesians interpret probabilities as degrees of belief, a certain amount of idealization is involved. The degrees of belief in question are not necessarily the actual degrees of conviction that a particular person has, but rather the degrees of belief that she would have if she were ideally rational and her degrees of belief were coherent.

Bayesians view the coherence requirement as having the same status and rationale as the requirement of logical consistency. Rationality requires not only consistency with the laws of logic, but also consistency with the probability axioms. Both are necessary conditions for rational belief. What makes the Bayesian position thoroughly subjective is its insistence that coherence (which entails logical consistency) is not only necessary but also sufficient for rationality: no matter how crazy one's degrees of belief may seem to someone else, if they satisfy the probability axioms, then, according to the Bayesians, they cannot be condemned as irrational.

Suppose that Adam and Eve each have a different degree of conviction in the proposition, Q , that it will snow in Phoenix next July. Adam gives it a probability of 0.9, while Eve gives it only a 0.05 chance of being true. As long as Adam also assigns a probability of 0.1 to the proposition that it will *not* snow in Phoenix next July his degrees of belief conform to the probability axioms. Similarly, if Eve's degrees of belief are coherent, then she thinks it 0.95 likely that there will be no July snowfall in Phoenix next year.

As far as subjective Bayesians are concerned, both Adam and Eve are rational with respect to the extremely limited set of propositions consisting of Q and $\sim Q$. But there is more to coherence than merely satisfying the negation rule. One also has to satisfy the special and general addition rules for disjunctions (p. 629), the special and general multiplication rules for conjunctions (p. 630), the implication and equivalence rules (p. 629), and axiom 2 concerning necessary truths (p. 628). Axiom 2 is especially problematic, since it requires that every necessary truth, no matter how complex, be assigned a probability of 1. Thus, while it might seem as though coherence is a very weak condition (because it places no restrictions on the degree of belief that a rational person can assign to any particular contingent proposition), in fact coherence makes very strong demands on the degrees of conviction that can be assigned to the members of any reasonably sized set of propositions (where that set includes many contingent statements, many necessary statements, and all their truth-functional compounds). Moreover, every proposition in the set must be assigned a precise number in the interval from 0 to 1.

Even when we have a coherent set of degrees of belief, there are several tricky issues connected with $P_{\text{new}}(T) = P(T/E)$, the Bayesian conditionalization rule that we are supposed to use in order to learn from experience. The first of these problems concerns the prior probability of T , $P(T)$, that appears on the right-hand side of the Bayesian expression for $P(T/E)$. Suppose one were to assign a prior probability of 0 to the theory,

T . In that case, no amount of evidence could ever confirm it, since its posterior probability, $P(T/E)$, would always remain 0. It is important to realize that the coherence requirement does not solve this problem. Coherence requires that we conform to the probability axioms. The relevant axiom (axiom 2) dictates that all necessary truths have a probability of 1. In conjunction with the other axioms, this entails that all necessary falsehoods must have a probability of 0. But the axioms do not forbid us from also assigning a probability of 0 to any contingent proposition that we might judge to be impossible. The usual Bayesian response is to impose the further demand of *strict coherence*: a set of beliefs is strictly coherent if and only if it is coherent and no contingent proposition is assigned a probability (degree of conviction) of 0 or 1. (This is one response to the Popperian argument discussed in the section "Why All Theories Are Improbable," in the commentary on chapter 1.)

Another problem with the dependence of $P(T/E)$ on $P(T)$ is that the amount by which evidence E confirms (or disconfirms) theory T will depend on the prior probabilities assigned to T by different scientists. How can this subjective influence on the degree of confirmation be reconciled with scientific objectivity? Bayesians reply by appealing to a theorem about the *washing out* (or *swamping*) of *priors*. The theorem shows that as evidence accumulates, the values of $P(T/E)$ calculated by individual scientists with different prior probabilities will tend to converge. In the long run, the initial divergence in the (subjective) prior probabilities becomes irrelevant and (objective) evidence dominates the calculation of confirmation.

Finally (for now), there is the issue of motivating the Bayesian conditionalization rule.⁶ Why should we revise our degrees of belief in accordance with Bayes's formula for $P(T/E)$? Remember that rationality, for Bayesians, is supposed to begin and end with the requirement of coherence (or strict coherence) for one's beliefs at any given time. Why, then, is it irrational for someone to violate the Bayesian rule when revising her degrees of belief, so long as her entire set of beliefs remains coherent? If a Bayesian were later to look at that coherent belief set, in ignorance of how the person's degrees of belief had been arrived at, the Bayesian would judge the set rational. The synchronic Dutch book argument (for the coherence of beliefs at a given time) seems to have no relevance to the diachronic conditionalization rule (for how probabilities should change over time). One popular Bayesian response to this challenge is to construct a further Dutch book argument that is explicitly diachronic.⁷ The gist of the argument runs as follows. If one makes a series of bets, some of which depend on what one's degrees of belief will be in the future, and if one follows a rule other than the Bayesian conditionalization rule and this alternative rule is known to the person with whom one is betting, then the person with whom one is betting can always construct a Dutch book against one. A necessary and sufficient condition for avoiding a Dutch book under the conditions stated is that one uses nothing but the Bayesian

conditionalization rule for changing one's degrees of belief. Just as the synchronic Dutch book argument is used to derive the coherence condition from the presumption of rationality, so, too, the diachronic Dutch book argument is supposed to show how rationality mandates the Bayesian rule for revising one's degrees of belief over time.

5.2 | Salmon on Kuhn and Bayes

In a companion article, written at the same time as the piece in our book, Salmon relates that, when he first began reading Kuhn's *The Structure of Scientific Revolutions* (1962), he was so shocked by Kuhn's repudiation of the distinction between the context of discovery and the context of justification that he set the book aside without finishing it.⁸ Later, in 1969, while preparing for a conference on the relation between the history of science and the philosophy of science, Salmon returned to Kuhn's book with renewed interest. Salmon conjectured that the reason for Kuhn's rejection of the traditional discovery-justification distinction lay in Kuhn's commitment to an inadequate conception of scientific justification, namely the H-D account. According to the H-D account, everything connected with the genesis of a scientific theory and its evaluation prior to being tested belongs to the context of discovery and, as such, is irrelevant to the theory's epistemic justification; a theory is justified and its acceptance rationally warranted only when the theory has been confirmed, and a theory is confirmed if and only if the predictions deduced from it are observed to be true.

THE INADEQUACIES OF HYPOTHETICO-DEDUCTIVISM

As Salmon notes, a significant limitation of the H-D account of confirmation is that it ignores statistical theories. Statistical theories confine their predictions to assignments of probability to classes of events but do not logically imply that any particular event will occur. By regarding inductive confirmation, in effect, as the inverse of logical deduction, the H-D account excludes from its scope all those theories in which the relation between theory and evidence is not deductive but probabilistic. The Bayesian approach has no such limitation, since it permits the likelihood $P(E/T)$ to assume values less than 1.

In its simplest form, the H-D account seems committed to the view that any theory that logically implies an observational prediction, O , is as well confirmed by that prediction as any other theory that implies O . This flies in the face of common sense and scientific practice. Bayes's equation is attractive because it can do justice to the differential confirmation of rival theories by the same evidence, by appealing to differences in the

initial plausibility of those theories (and hence differences in their prior probabilities). A nice illustration of this is the Bayesian solution to the so-called tacking paradox or the problem of irrelevant conjunction. Suppose that theory T , in conjunction with background information B , entails the true observational prediction E : $(T \& B) \rightarrow E$. Now let I stand for some contingent statement that is logically independent of T and irrelevant to E . It follows trivially that if $(T \& B) \rightarrow E$, then it must also be the case that $(T \& I \& B) \rightarrow E$. Thus, according to the H-D account, E confirms both the original theory T and the augmented theory $(T \& I)$ and, moreover, confirms them by the same amount. The Bayesian analysis agrees with the H-D account that E confirms both theories but disagrees about the degree of confirmation. On the most popular version of the Bayesian analysis, the degree of confirmation of a theory by evidence is a function of the difference between the posterior probability of the theory given that evidence and the prior probability of the theory.⁹ On this version, the Bayesian analysis entails that the degree of confirmation conferred by E on $(T \& I)$ must be less than the degree of confirmation that E confers on T alone. Here are the two expressions for the degrees of confirmation of T and $(T \& I)$ on the difference analysis.

$$P(T/E \& B) - P(T/B) = P(T/B) \times \left[\frac{1 - P(E/B)}{P(E/B)} \right]$$

$$P(T \& I/E \& B) - P(T \& I/B) = P(T \& I/B) \times \left[\frac{1 - P(E/B)}{P(E/B)} \right]$$

These expressions are derived from the second version of Bayes's equation, setting $P(E/T \& B)$ and $P(E/T \& I \& B)$ both equal to 1. The factor in the square brackets is the same for both theories, and their respective degrees of confirmation are proportional to their prior probabilities. Since $(T \& I) \rightarrow T$, and I is a contingent statement that is independent of T , it follows from the implication rule that $P(T \& I/B)$ must be less than $P(T/B)$. Thus, E confirms $(T \& I)$ by a smaller amount than it confirms T : adding the irrelevant conjunct I to T lowers the confirmation provided by E . In this respect, then, the Bayesian approach to confirmation is a decided improvement over the H-D account.

Although Salmon does not discuss it in his article, the Bayesian approach also promises a resolution of the Duhem problem (that is, the problem of assigning the blame for a failed prediction to a particular member of a group of hypotheses), for not only are some theories confirmed better by the same evidence, but Bayes's equation can also be used to explain how some components of a group of hypotheses and assumptions receive a much larger *disconfirmation* than other components when observations disagree with theoretical predictions.¹⁰ Thus, the Bayesian ap-

proach can show us where the blame should be placed when a group of hypotheses and assumptions lead to a false prediction.

Salmon is not alone in thinking that the H-D account is inadequate. An early critic of hypothetico-deductivism, Popper rejected it because he denied the whole notion of inductive confirmation. (See Popper, "The Problem of Induction," in chapter 4 for details.) Other critics, such as Carnap and Reichenbach, accepted that confirmation is an essential part of scientific rationality but insisted that it should be understood in terms of Bayes's theorem. Carnap interpreted the probabilities in Bayes's theorem as a priori logical probabilities; Reichenbach construed them as empirical frequencies. Both were objectivists about probability. More recently, an entire school of statisticians and philosophers of science has arisen—the personalists or Bayesians—that interprets the probabilities in Bayes's theorem subjectively, as degrees of belief.

Salmon thinks that we can use Bayes's theorem to reconcile Kuhn's historical approach to understanding science with the logical empiricism of philosophers such as Carnap, Reichenbach, and Hempel. The key is to incorporate Kuhn's values—criteria for theory assessment such as consistency, simplicity, and fruitfulness—into the Bayesian equation that defines confirmation. Variation in the interpretation of these values and the emphasis placed on them can give rise to differing judgments about the prior probability of a theory. Thus, scientists can reach different but equally rational judgments about how well a theory is confirmed by a particular piece of evidence using the same algorithm (Bayes's equation) because they insert different inputs into that algorithm in the form of different judgments about prior probability. Salmon contrasts this with Kuhn's own suggestion in "Objectivity, Value Judgment, and Theory Choice" (reprinted in chapter 2, above) that scientists reach different conclusions because they use different algorithms. However, careful study of Kuhn's article reveals that when Kuhn talks about using different algorithms he really means inserting different subjective inputs into a Bayesian algorithm, so Kuhn is much closer to Salmon's position than his language might suggest.

PRIOR PROBABILITIES

As Salmon emphasizes in his article, most of what is philosophically controversial about the Bayesian approach to confirmation depends on the interpretation of prior probabilities in Bayes's equation. Salmon distinguishes three such interpretations: the objective-logical, the objective-empirical, and the subjective. Salmon agrees that the objective-logical interpretation of Carnap and others, according to which probabilities are assigned a priori to all statements on the basis of a formal language and assumptions about the equiprobability of states of affairs, is hopelessly inadequate to the task of analyzing the probability of real-life scientific the-

ories. That leaves the objective-empirical and the subjective interpretation.

In an earlier book, *The Foundations of Scientific Inference*, Salmon adopted Reichenbach's objective-empirical interpretation according to which probabilities are relative frequencies.¹¹ The basic idea is this: Every hypothesis is either true or false, but when a new hypothesis, H , is first proposed, we do not know which attribute (truth or falsity) it has. In order to make sense of the prior probability (or plausibility) of the single hypothesis, H , we have to place it in a reference class of similar hypotheses. Then, on the basis of past experience, we can see how often hypotheses in this class have turned out to be true. The ratio of the number of true hypotheses to the total number of hypotheses in the reference class is then taken to be the prior probability of H .

In rough outline, this procedure is supposed to be similar to the way frequency theorists handle the problem of the single case. A typical example is the problem of assigning a probability to whether a particular (asymmetrical) coin will show a head on its next toss. In the case of the coin, we estimate this probability by dividing the number of times the coin has come up heads by the number of times the coin has been tossed. Thus, very roughly, the frequency theorist would say that the probability of getting a head on the next toss of the coin is 0.55 if the frequency of heads converges to 0.55 as the number of tosses becomes ever larger. Applying this same approach to the prior probability of hypotheses is extremely difficult. Not least among these difficulties is the problem of specifying the appropriate reference class. What exactly does it mean to talk about hypotheses that are similar to H ? Is it a matter of mathematical form, such as the use of inverse-square laws? And if so, why should the success of such laws in one domain of science (such as the study of gravity) make it more likely that they will succeed in another domain (such as the investigation of the strong force binding together particles in the atomic nucleus).¹²

Throughout his career, Salmon has been highly critical of the unfettered subjectivism of the pure Bayesian or personalist interpretation of probabilities as degrees of belief. To Salmon, scientific judgments about confirmation should not depend in any way on the prejudices, emotions, or mood swings of individual researchers. The answer, he thinks, lies in what he (following Abner Shimony) calls *tempered personalism*. Tempered personalism places constraints on prior probabilities that go beyond mere coherence. Since experience has taught us that scientists have been moderately successful in the past, no hypothesis advanced by a serious scientist should be given a prior probability that is either 0 or vanishingly small. But, again, since experience tells us even the most promising hypotheses in the past have sometimes turned out to be false, the prior probability of any new hypothesis should be fairly low. The notion of *success* invoked in this discussion is crucial for understanding Salmon's proposal. For Salmon believes that when we assign a prior probability to a new hypothesis, we

are trying to estimate the correct, objective probability that the hypothesis will turn out to be successful, and by *successful*, Salmon means *true*. Thus, while it may seem as if Salmon is making significant concessions to Kuhn when he admits consistency, analogy, and professional scientific standing as factors that play a legitimate role in determining the plausibility of new hypotheses, in fact his conception of probability is still, at bottom, objective and frequentist. It is, for example, only because past experience has taught us that hypotheses advanced by cranks very seldom turn out to be true that we should assign them a negligibly small prior probability. As Salmon himself puts it, “prior probabilities . . . can be understood as our best estimates of the frequencies with which certain kinds of hypotheses succeed. . . . The personalist and the frequentist need not be in any serious disagreement over the construal of prior probabilities” (564).

One final but important point: Salmon readily admits that it “seems preposterous” (564) that plausibility judgments based on values such as simplicity and symmetry could result in exact numbers for prior probabilities. Like many advocates of Bayes’s equation, Salmon appeals to the washing out or swamping of priors to argue that their exact value really does not matter. For as soon as evidence begins to accumulate, the values for the posterior probability of a hypothesis converge. In the long run, the particular values adopted for the prior probability become irrelevant (so long as we avoid the extreme values of 0 and 1). But this convergence argument assumes that different scientists agree on the likelihoods, an assumption that Salmon defends later in his article.

THE EXPECTEDNESS OF EVIDENCE

Methodologists of science commonly hold that a theory receives greater confirmation from the successful prediction of something surprising than from the prediction of something expected. This issue is addressed, in part, in chapter 4 under the guise of the debate over novel predictions. Because the right-hand side of Bayes’s equation has $P(E)$, the probability of the evidence E , as its denominator, it follows that, other things being equal, the lower the value of $P(E)$, the greater the value of $P(H/E)$. Thus, the more unexpected the prediction, the greater its confirming power if it should turn out to be true. But what is $P(E)$, the expectedness of the evidence, and how can it be measured?

Salmon uses the total probability rule to express $P(E)$ in terms of prior probabilities and likelihoods, writing all the probabilities involved as conditional on background knowledge B , where B includes initial conditions, boundary conditions, auxiliary hypotheses, and other relevant information.

$$P(E/B) = P(E/T \& B) \times P(T/B) + P(E/\sim T \& B) \times P(\sim T/B).$$

If T is a deterministic theory, then T (in conjunction with auxiliary hypotheses and assumptions) entails E . In such a case, the likelihood $P(E/T \& B)$ equals 1 and $P(E/B)$, the expectedness of the evidence, must be at least as great as $P(T/B)$, the prior probability of T . But assigning an exact number to $P(E)$ is not easy, since it involves knowing the value of the likelihood $P(E/\sim T \& B)$, a problem that Salmon addresses in his section on likelihoods.

A second difficulty with $P(E/B)$ that Salmon acknowledges is a version of the problem of old evidence. This is discussed at some length later in this commentary in the section "The Problem of Old Evidence." For the moment we merely note that Salmon thinks that, given his characterization of background information B , the objective value of $P(E/B)$ must always be 1. Since B includes all the details about the experimental setup and the instruments used to observe E , the objective probability that B will occur under those conditions (assuming that the system in question is deterministic) is 1. Thus, Salmon concludes that the expectedness $P(E)$ can only be a subjective probability, reflecting the degree to which a particular scientist finds E psychologically surprising. Given Salmon's hostility towards subjectivism, the conclusion that "expectedness defies interpretation as an objective probability" (566) is highly unwelcome. At the end of his article, Salmon suggests a way to avoid this and a similar problem with the likelihood $P(E/B \& \sim T)$, while still permitting objective comparisons among rival theories.

LIKELIHOODS AND THE CATCH-ALL HYPOTHESIS

The main problem with likelihoods concerns the value of $P(E/\sim T \& B)$, which appears in the expression for $P(E)$. $P(E/\sim T \& B)$ is the probability that E is true given that theory T is false, but since $\sim T$ is not a specific theory, the corresponding likelihood is not well defined. Even when we have two competing theories, T_1 and T_2 (such as specific versions of the wave and particle theories of light), $P(E/\sim T_1)$ is not equal to $P(E/T_2)$. Although theories T_1 and T_2 are contraries, and thus T_1 entails $\sim T_2$ and T_2 entails $\sim T_1$, they are not contradictories; thus, $\sim T_1$ does not entail T_2 , nor is T_1 logically equivalent to $\sim T_2$. It is possible that both T_1 and T_2 are false. Thus, if we write out the set of logically exclusive and exhaustive hypotheses, it will include not only T_1 and T_2 , but also T_k , the so-called catch-all hypothesis. What is the catch-all hypothesis? Strictly speaking, it is not a single hypothesis at all but a lengthy disjunction of all the possible alternatives to T_1 and T_2 , most of which we have never thought about. As Salmon says, trying to guess the ingredients of the catch-all would be like trying to predict the future of science. Even though some of these ingredient hypotheses entail E , this scarcely helps us to answer the question, What is $P(E/T_k)$, the likelihood of T_k ? because T_k is the disjunction of *all*

the possible alternatives, including those that do *not* entail E . And if we cannot answer this question, then we cannot calculate $P(E)$. Because Salmon regards the problem of calculating the likelihood of the catch-all as completely intractable, he proposes a method for choosing between theories that does not require the calculation of $P(E)$.

SALMON'S BAYESIAN ALGORITHM FOR THEORY PREFERENCE

Salmon agrees with Kuhn that theory choice in science is usually a comparative affair. Typically, the issue is not how well a particular piece of evidence confirms an individual theory, but how well that evidence favors one theory over its rivals. In any given scientific domain only a few theories—usually just two or three—will be competing for acceptance at any given time. Certainly the catch-all hypothesis is seldom a serious option. Thus, despite the intractability of calculating the likelihood of the catch-all and the expectedness of the evidence, the Bayesian approach can still reflect the realities of scientific practice if it can provide a comparative ranking of those hypotheses that are serious rivals. Salmon's proposal is that, in choosing between two theories, T_1 and T_2 , on the basis of evidence E , we should compare the posterior probabilities $P(T_1/E\&B)$ and $P(T_2/E\&B)$. An attractive feature of this proposal is that, in forming the ratio of the posterior probabilities, the problematic term $P(E/B)$ cancels out.

$$\frac{P(T_1/E\&B)}{P(T_2/E\&B)} = \frac{P(E/T_1\&B) \times P(T_1/B)}{P(E/T_2\&B) \times P(T_2/B)}.$$

Assuming that T_1 and T_2 are the only candidates for serious consideration, Salmon's proposal is that, before the discovery of evidence E , scientists should prefer T_1 to T_2 if and only if the prior probability of T_1 is greater than the prior probability of T_2 . After the discovery of E , scientists should change their preference from T_1 to T_2 if and only if the posterior probability of T_2 is greater than the posterior probability of T_1 . It follows from the Bayesian expression for the ratio of the posterior probabilities that, after the discovery of E , scientists should prefer T_2 to T_1 , if and only if

$$\frac{P(E/T_2\&B)}{P(E/T_1\&B)} > \frac{P(T_1/B)}{P(T_2/B)}$$

or, in other words, if and only if the ratio of the likelihoods is greater than the reciprocal of the ratios of the prior probabilities. Salmon refers to this as the *Bayesian algorithm for theory preference*.

Salmon's algorithm is both ingenious and attractive, but it also has its limitations and counterintuitive features. First, it should be clear that in "choosing" T_2 over T_1 , we are not deciding to accept T_2 as true or well-

confirmed. We are merely saying that, relative to evidence E , T_2 is better confirmed than T_1 . For all we know, T_2 might be extremely improbable and unworthy of acceptance. It is important to remember that, in comparing the posterior probabilities of T_1 and T_2 , we are not calculating—nor, if Salmon's pessimism is correct, can we ever calculate—the degree of confirmation of either hypothesis.¹³ Thus, the judgment resulting from Salmon's algorithm is relatively weak, since it merely asserts that evidence E supports one theory better than its rival. The degree of that support is left entirely undetermined.¹⁴

Second (as noted by Wade Savage, the editor of the volume in which Salmon's article first appeared) Salmon's algorithm cannot, in its present form, give us any rational guidance when, as often happens, the body of evidence for T_1 is different from the body of evidence for T_2 . For, obviously, $P(E)$ cancels out only when the evidence, E , is the same for both theories. Similarly, Salmon's algorithm does not permit us to judge whether one piece of evidence confirms a theory better than another piece of evidence.

Third, as Salmon himself notes, when both theories are deterministic and, in conjunction with B , entail the evidence E , the ratio of their posterior probabilities given E reduces to the ratio of their prior probabilities. Thus, according to Salmon's algorithm, no amount of evidence can change our initial preference ranking for such theories. For deterministic theories, the likelihoods become irrelevant and the prior probabilities (influenced by Kuhn's criteria for theory choice) dominate completely. Anyone who is critical of the vagueness of Kuhn's criteria and the difficulty of weighing and comparing them is unlikely to be impressed by this as a demonstration of a rational algorithm underlying scientific decisions about theories.

Salmon's response to this third point is contained in sections 8 and 9 of his paper. In section 9, "Kuhn's Criteria," Salmon distinguishes three types of theoretical virtue: informational, economic, and confirmational. Salmon argues that two of Kuhn's criteria—scope and accuracy—fall outside the confirmational category and are thus irrelevant to the prior plausibility of theories. This reduces the task of making the basis of plausibility judgments more precise by narrowing the focus to Kuhn's remaining three criteria, namely, simplicity, consistency, and fruitfulness.

In section 8, "Plausible Scenarios," Salmon explains that, when they are first formulated, important scientific theories often have great difficulty in explaining some puzzling phenomenon. He gives as examples the difficulty the absence of detectable stellar parallax posed for the Copernican theory and the problems of giving a coherent account of the optical ether and the phenomenon of selective absorption faced by the wave theory of light. Salmon's point is that the original versions of these theories did not logically entail the phenomena they had difficulty explaining. Indeed, the probability of E (a puzzling phenomenon) given T (the theory in question)

was rather low. The challenge for supporters of T was to come up with what Salmon calls a *plausible scenario*, that is, a set of further assumptions, A , that when added to T , would significantly raise the probability of E . Ideally, when the augmented theory ($T\&A$) entails E , the likelihood $P(E/T\&A\&B)$ attains its maximum value of 1. Then, applying Salmon's algorithm to choose between the augmented version of T and its rivals, the prior probabilities of ($T\&A$) and its competitors become crucial. This is where Kuhn's criteria of simplicity, consistency, and fruitfulness for assessing plausibility become important. Thus, when T is first proposed, likelihoods are not irrelevant because their low values motivate the search for plausible scenarios. If that search is successful, then likelihoods become irrelevant because judgments of prior probability are made by applying plausibility criteria, not to the original theory T , but to the new augmented version ($T\&A$). So the allegation that, according to Salmon's algorithm, theory preference is permanently determined by the initial prior probabilities ignores the dynamics of scientific research.

5.3 | Why Glymour Is Not a Bayesian

Glymour's criticisms of Bayesianism fall into three main categories. First is Glymour's attack on the a priori arguments that Bayesians give to justify coherence, conditionalization, and the other constraints they impose on belief and belief change. As Glymour notes, his criticisms of these arguments are not original, and Bayesians themselves admit that their attempts at justification fall short of being conclusive. Second are Glymour's reasons for judging as inadequate Bayesian attempts to explain widely accepted views about confirmation, such as scientists' preference for variety in evidence and simplicity in theories, and their general disdain for ad hoc hypotheses. The emphasis on explanation is important. Glymour contends that it is not enough for Bayesians merely to give a framework that can accommodate these methodological truisms by placing restrictions on prior probabilities. Rather, the Bayesians must give a coherent and plausible rationale for these restrictions that explains why they should hold. This, Glymour claims, the Bayesians have failed to do. Third are Glymour's criticisms of Bayesianism for its alleged inability to shed light on evidential relevance. When does a piece of evidence confirm a theory? Why does some evidence confirm some parts of a theory but not others? To these questions, Glymour judges that Bayesians have given no satisfactory answer.

Glymour regards explicating the concept of evidential relevance as a crucial problem for the philosophy of science. In the rest of his book, *Theory and Evidence*, from which our reading in chapter 5 is taken, Glymour defends his own bootstrap theory of confirmation. Glymour claims

that his theory analyzes evidential relevance in a way that accords with actual scientific practice and that shows how some pieces of evidence can confirm some parts of a theory but not others. In this way, Glymour thinks we can avoid the horror of Quinean holism—the doctrine that theories can be tested and confirmed only as unanalyzable wholes—without relying on the analytic-synthetic distinction. (For a discussion of Quinean holism and the analytic-synthetic distinction, see the readings by Quine, Gillies, and Laudan in chapter 3 and the accompanying commentary.)

BAYESIAN A PRIORI ARGUMENTS

Glymour concedes that the Bayesian appeal to betting behavior does provide evidence that people have degrees of belief, at least with regard to propositions on which they are willing to bet, and that these degrees of belief can be measured, at least approximately, on a scale from 0 to 1. But the argument from betting behavior is not conclusive, since someone can wager at a particular set of odds without those odds reflecting the bettor's degree of belief in the proposition that is the object of the wager.

Even if there are degrees of belief, why must they conform to the probability axioms as the Bayesians require when they impose the condition of coherence? As discussed earlier in this commentary, the standard Bayesian argument for coherence is the Dutch book argument. One key premise in this argument is that it would be irrational for a person to accept a sequence of bets such that, no matter how things turn out, she is bound to lose money. One way to avoid a Dutch book is to ensure that one's degrees of belief (and, hence, for the Bayesian, the least odds at which one is willing to bet on various propositions) obey the probability axioms. But that is not the only way avoid the certainty of losing money. Another way is to refuse to place any group of bets that would result in a Dutch book. For example, Samantha might be willing to bet on P at odds of 3 to 1 and on $(P \& Q)$ at odds of 5 to 1, but as long as she refuses to place *both* bets at the same time, no Dutch book against her will result. Indeed, rational bookies try to guarantee that they will make money, no matter how events turn out, by deliberately avoiding coherence in the odds they offer gamblers. Thus, coherence does not follow deductively from the avoidance of a Dutch book or from rationality in one's betting behavior.¹⁵

What then of the other Bayesian constraint, namely, the use of Bayes's equation to revise probabilities in the light of evidence, E , by replacing one's old prior probability, $P(T)$, with the posterior probability, $P(T/E)$? As explained earlier (pp. 637–38), the conditionalization rule for revising degrees of belief over time does not follow from the requirement that, at all times, one's degrees of belief be coherent. Some authors (Paul Teller, David Lewis) have offered a diachronic Dutch book argument for using conditionalization to revise degrees of belief over time. But, like the synchronic Dutch book argument for coherence, it falls short of being a proof.

At best it shows that if you eschew conditionalization, then you would be irrational to bet honestly with someone who knows your strategy for changing your betting quotients.

Another argument for conditionalization that Glymour discusses is based on the phenomenon of “washing out” or “swamping” of prior probabilities (see pp. 590–91). Suppose that one is persuaded by Bayesians that a person’s degrees of belief at a given time should be coherent and that there are no other rational constraints on prior probabilities. In that case, the prior probability assigned to a given theory is entirely subjective and, despite coherence, there is not the least reason to expect that it will reflect the true, objective probability of the theory in question. Given this subjectivity of prior probabilities, why should one use Bayesian conditionalization (rather than, say, random guessing or consulting an oracle) to revise one’s beliefs in response to new evidence? Why should a posterior probability derived by applying Bayes’s equation to a purely subjective prior probability be any more likely to reflect the truth than a new prior probability generated in some other way? The Bayesian answer is as follows: conditionalization is rational because, if people use conditionalization to revise their beliefs, then, in the long run, no matter how different their initial priors, they will come to agree on which theory among the competing alternatives is true (or overwhelmingly probable). In a nutshell, and subject to several conditions, conditionalization guarantees convergence to the truth.

Now, as Glymour points out (echoing the criticisms of Mary Hesse), the relevance to scientific inference of the convergence theorem (or, as Glymour calls it, the *stable estimation theorem*) proved by Leonard Savage is dubious. Before explaining these doubts, let us take a closer look at Savage’s result. Since Glymour’s notation (following Savage) is rather formidable, we have simplified it slightly (relying heavily on Hesse’s discussion). We start with the assumption that we have a finite set, B , of q hypotheses that are mutually exclusive and exhaustive. In other words we have hypotheses, B_1, B_2, \dots, B_q , such that it must be the case that exactly one of them is true and the rest false. Assume that the true hypothesis is, in fact, the first one, B_1 . We have some body of evidence, x , the outcome of a series of measurements or experiments, that consists of a specific sequence of values for the first n of the variables x_i . Each of the hypotheses, B_i , assigns a different probability to x . More generally, in Glymour’s language, for every possible sequence of the n specific values that could have made up x , each hypothesis assigns a different probability to that sequence of values; each hypothesis yields a different likelihood distribution.

A simple illustration of these conditions is provided by the case of an urn containing eighty balls at least one of which is black (B). Relative to this information, there are eighty mutually exclusive and exhaustive hypotheses— B_1 : exactly one ball is black, B_2 : exactly two balls are black,

and so on. After mixing the balls thoroughly, we draw a single ball from the urn, observe its color, and then replace it. The result of our observation is the value of the first variable, x_1 . We repeat this process of random sampling with replacement ten times, thus generating the body of evidence, x . The sequence observed is, let us assume, $\sim B, B, \sim B, \sim B, B, \sim B, B, B, B, B$. Each observation is independent of every other, and each hypothesis yields a different probability for a sequence in which three-fifths of the ten balls drawn are black. If, indeed, the true hypothesis is B_1 and there is only one black ball in the urn, then the likelihood $P(x/B_1)$ is very low and can be calculated, as can all the other likelihoods, using the binomial theorem.¹⁶

Now consider the Bayesian expression, $P(B_1/x)$, for the posterior probability of hypothesis B_1 given evidence x . Clearly, this will be something other than 0 only if we impose the further condition that, for each hypothesis, its prior probability is greater than 0. With that condition in place, Savage then proves the following result about the second-order probability of the true hypothesis B_1 —the probability that $P(B_1/x) > \alpha$, for any fixed number α that can be as close to 1 as we please. If we use Bayesian conditionalization and if the number of observations, n , included in x is sufficiently large, then the probability that $P(B_1/x) > \alpha$ approaches 1 as n increases. In plain English, as the number of observations increases, the posterior probability of the true hypothesis is increasingly likely to converge to 1, and the posterior probability of each of the false hypotheses is increasingly likely to converge to 0.

Glymour makes two criticisms of Savage's theorem as an attempt to vindicate Bayesian conditionalization. First, he contends that the theorem does not actually say that as evidence increases the posterior probability of the true hypothesis will approach 1. Rather, the theorem makes a second-order claim about the probability that this probability will approach 1. Thus, if the Bayesian is consistent and interprets all probabilities as degrees of belief, then the theorem tells us that, *if we are Bayesians*, then our degree of belief will be extremely high that we will end up assigning a high probability to true hypotheses. However reassuring this may be for Bayesians, Glymour doubts its significance for non-Bayesians (like himself) who are seeking a reason to convert to Bayesianism. In response, Bayesians could say that insofar as the original challenge was to justify conditionalization within the resources of the Bayesian theory, the theorem succeeds. The issue of whether the theorem also provides a convincing reason for non-Bayesians to become Bayesians is a separate issue.

Second, Glymour argues that several conditions required for Savage's proof make his theorem inapplicable to real cases of scientific inference. Principal among them are the assumptions that (i) we start with a set of hypotheses that are logically exclusive and exhaustive; (ii) each of these hypotheses yields a well-defined probability for the evidence under consideration, and no two hypotheses yield the same probability; (iii) the order

in which the evidence is collected is irrelevant: the observations are a simple, random selection from all the observations that could have been made; (iv) for every hypothesis, each observation is statistically independent of every other. In short, the conditions that make sampling without replacement from an urn a good illustration of Savage's theorem are conditions that scientific theories never satisfy.

METHODOLOGICAL TRUISMS

As Glymour notes, few Bayesians regard the *a priori* arguments as convincing. Instead, they rest their case largely on the presumed ability of Bayesianism to do justice to widely accepted principles of scientific reasoning. Here Glymour demurs. He concedes that any particular inference can be made consistent with the Bayesian apparatus by adjusting degrees of belief. But, he objects, this sort of gerrymandered agreement falls short of explaining why the inference in question is good. An inference is an argument. Thus, to explain in a philosophically insightful way why it is good means showing why any rational person who accepts its premises should also accept its conclusion. Merely finding an assignment of prior probabilities that makes plausible a particular person's transition from believing the premises to believing the conclusion does not explain why the argument warrants belief in its conclusion. The Bayesian account ignores the most distinctive feature of arguments, namely their use as impersonal instruments for rational persuasion.

Glymour discusses two methodological truisms: our preference for variety in evidence and simplicity in theories. His discussion of evidential variety is brief. Glymour has no doubt that the Bayesian framework can accommodate our preference for diverse evidence. When comparing rival theories, for example, diverse evidence is preferable because it is more likely to reveal which of the competitors are false. (See "Diversity of Evidence" later in this commentary for a discussion of a Bayesian argument along these lines by Paul Horwich.) But as Glymour sees it, there is one important rationale for evidential variety that the Bayesian approach is unable to explicate, namely our desire to test (and, hence, confirm) the independent parts of complex theories. In general, particular pieces of evidence will be relevant to some parts of a theory but not to others. Diverse evidence tests more of the parts of a complex theory than does narrow evidence, and by testing more of the theory, diverse evidence is able to confirm it more strongly. The key to this rationale is evidential relevance, and the fact that the consequences derived from a theory differ in their confirmational power. Glymour takes up this topic towards the end of his reading. (See "The Relevance of Evidence to Theory" later in this commentary.) For the moment, he turns his attention to the other methodological truism, our preference for theoretical simplicity.

SIMPLICITY

The preference for simplicity manifests itself in various ways. First, it is reflected in our disdain for what Glymour calls *deoccamized* theories. *Deoccamization*, a term that Glymour borrowed from David Kaplan, is the antithesis of Ockham's razor.¹⁷ To deoccamize a theory, replace one of its theoretical terms with a fixed combination of new terms. For example, using terms invented by Glymour, we might replace every occurrence of *force* in Newtonian mechanics with the sum of *gorce* and *morce*.¹⁸ We have no further information about *gorce* and *morce* and no way of measuring either of them as independent quantities, but because of the way the replacement has been carried out, the deoccamized *gorce-plus-morce* theory has exactly the same observational consequences as the Newtonian theory. Since both theories entail the same evidence, their likelihoods are the same: for any evidence, *E*, the probability of *E* given the Newtonian theory is exactly the same as the probability of *E* given the *gorce-plus-morce* theory. Thus, the Bayesian cannot appeal to any difference in likelihoods to explain why the Newtonian theory is much better confirmed than its deoccamized cousin.¹⁹ The only other way the Bayesian can explain a difference in confirmation lies in the prior probabilities. But, Glymour argues, being deoccamized, having "extra wheels" as he puts it, cannot explain why a theory has a low prior probability, because whether a theory has "extra wheels" can only be judged relative to a body of evidence. By restricting the body of evidence, however artificially, just about any theory, even a scientifically respectable theory such as Newton's, would be judged to be deoccamized. Thus all theories would qualify as being deoccamized and have to be assigned the same low prior probability.²⁰

A second way in which the preference for simplicity manifests itself is in regard to curve fitting. The received wisdom is that the simplest equation that entails the data is the best choice and the best confirmed. (See, for example, the curve-fitting diagram in Paul Horwich's article "Wittgensteinian Bayesianism," which follows the reading from Glymour in this chapter.) In order to explain this methodological truism, the Bayesian has two choices: either define a measure of simplicity for equations and argue that simpler equations have higher prior probability or, failing that, argue that simpler theories have higher likelihoods. Glymour examines two instances of the first strategy, by Harold Jeffreys and Mary Hesse, and argues that neither succeeds. He also evaluates Roger Rosenkrantz's version of the second strategy and argues that it, too, fails.

Jeffreys labels the idea that simpler theories have higher prior probability, the *simplicity postulate*. He proposes that the simplicity of a theory depends on the number of independent parameters in the equations it uses: the fewer the parameters, the simpler the theory. But Jeffreys' pro-

positional founders because it entails that hypotheses having the same number of independent parameters have the same degree of simplicity and hence the same prior probability. Unfortunately, at any level of simplicity thus defined, the number of hypotheses is infinite, so the prior probability of every hypothesis would be 0. More importantly, as Glymour notes, even if finite probabilities could be assigned to the infinite number of hypotheses at each level, Jeffreys' approach is fundamentally at odds with our scientific practice. Assignments of prior probability are seldom, if ever, based solely on parameter counting. No one, for example, would suggest that Boyle's law is more likely to be true than van der Waals's equation simply because it involves fewer parameters. Indeed, given the frequency with which very simple numerical equations have turned out to be false, we have every reason to assign low prior probabilities to the simplest equations. In actuality, judgments of prior probability are based on a complex mix of plausibility considerations in which parameter counting plays a minor, often negligible, role.

There is another, fatal flaw with Jeffreys' proposal. As Malcolm Forster has pointed out, hypotheses with few parameters trivially entail hypotheses with more parameters.²¹ For example, if all planets move in circles, then it follows that all planets move in ellipses. Why? Because a circle is a degenerate case of an ellipse in which the major and minor axes have the same length (namely, the radius of the circle). Similarly, hypotheses involving straight lines are special cases of, and hence imply, hypotheses using parabolas and higher-order curves. The implication rule tells us that when *A* logically entails *B*, the probability of *B* must be at least as great as the probability of *A*. Thus, the simplicity postulate is violated: simple hypotheses must be less probable, not more probable, than complex ones, if simplicity depends solely on parameter counting.

Hesse approaches simplicity—and, more generally, inductive inference—by means of a clustering postulate, a general assumption about the world that Hesse adds to the Bayesian machinery. In broad outline, the clustering postulate says that the world is more likely to be homogeneous than heterogeneous. More specifically, it says that instances tend to be clustered in the sense that $r + 1$ positive instances are initially more probable than r positive and one negative.²² Along the same lines, clustering entails that for any given r instances of property *P*, it is initially more probable that all or none of them will also be instances of property *Q* than that they will manifest any other proportion of *Q*s. Hesse then claims that this clustering postulate can explain our preference for simpler (that is, more economical) theories. One peculiarity of Hesse's discussion should be noted at this stage. She is not claiming that, given an observation of r individuals with both *P* and *Q* (and no observations of individuals that have *P* but lack *Q*) that the theory "All *P*s are *Q*" has a high prior probability. Indeed, she agrees with Popper that, when applied to infinite domains, the probability of any universal theory is 0. Rather, her claim is

that "All P s are Q " is more likely to be true than its rivals when applied to a finite range of cases, such as predictions about the next instance of P . In Hesse's view, all inductive inference is really from particulars to other particulars.

In the curve-fitting example discussed by Hesse and Glymour, the entirety of our evidence consists of the two data points (x_1, y_1) and (x_2, y_2) . Under those circumstances, Hesse argues that the best, most economical hypothesis is the linear equation $y = a + bx$. Her reasoning is as follows. Consider a third, as yet unexamined point g , whose x -value is x_3 . Given the clustering postulate, the most probable y -value for g is $a + bx_3$, which corresponds to choosing the "simplest" hypothesis $y = a + bx$. Other, higher-order equations are consistent with the data, but, she claims, only the linear equation has the values of all its parameters determined by that evidence. Thus, the clustering postulate requires that we regard the two observed data points as instances of the hypothesis $y = ax + b$. Unfortunately, as Glymour demonstrates, the appearance of a unique choice here is an illusion. As far as the clustering postulate is concerned, we could just as well regard the two observed data points as instances of the quadratic equation $y = a_1 + b_1x^2$, where the parameters a_1 and b_1 are defined by Glymour in his article, and their values, too, are fully determined by the existing evidence, since they are functions of x_1, y_1, x_2 , and y_2 . Obviously, we cannot assign a prior probability greater than $1/2$ to both the linear hypothesis and the quadratic hypothesis, since they are logical contraries. So Glymour judges Hesse's attempt to explicate simplicity within a Bayesian framework a failure.

Rosenkrantz's proposal is complicated, mainly because his second strategy (relating simplicity to likelihood) must be applied to theories that do not deductively entail the evidence. (As we have already seen, any theory that entails its evidence automatically gets the maximum likelihood of 1, thus ruling out likelihood as a way of discriminating among such theories.) So the issue is determining how well the evidence fits a number of theories, none of which entails that evidence. Statisticians use several criteria of goodness of fit. Without fussing over the technical details, suppose that we adopt one such criterion and suppose further that we are dealing with the simplest kind of case in which the measuring process that generates the evidence can have only a finite number of outcomes. In that case, Rosenkrantz defines what he calls the *observed sample coverage of a theory* as the number of possible outcomes that satisfy the goodness of fit criterion for that theory divided by the number that do not. Rosenkrantz proposes that the smaller the observed sample coverage, the simpler the theory. Moreover, he proves that when one hypothesis, H_1 , can be derived from another, H_2 , by holding constant one of the free parameters in H_2 , then the average likelihood of H_1 is greater than the average likelihood of H_2 . (We need not worry about how, exactly, *average likelihood* is defined for these hypotheses.²³) Yet because of the way H_1 is

derived from H_2 , its observed sample coverage is smaller than the observed sample coverage of H_2 , and hence H_1 is the simpler hypothesis. In this way, Rosenkrantz claims to have given a Bayesian vindication for our preference for simpler theories. Other things being equal, simpler theories are more strongly confirmed by the evidence because they have higher (average) likelihoods.

Glymour has two criticisms of Rosenkrantz. First, observed sample coverage is a poor measure of simplicity because, intuitively, deoccamized theories and theories with irrelevant conjuncts tacked onto them are less simple than the theories from which they were derived. But on Rosenkrantz's proposal, they have the same observed sample coverage. Second, the theorem proved by Rosenkrantz is too limited in its scope. It applies only to pairs of hypotheses, H_1 and H_2 , in which the simpler hypothesis, H_1 , is derived from the more complex hypothesis, H_2 , by fixing the value of one of the free parameters in H_2 . But comparisons of simplicity extend more widely than this. For example, any quadratic equation (with three free parameters) is less simple than any linear equation (with two free parameters), regardless of whether the latter can be derived from the former in the manner required by Rosenkrantz's theorem. Thus, Glymour complains, Rosenkrantz has not proven that, in general, simpler hypotheses have higher average likelihoods.²⁴

THE RELEVANCE OF EVIDENCE TO THEORY

Glymour raises three difficulties for any Bayesian account of the relevance of evidence to theory. Because of its importance and complexity, the last of these—the problem of old evidence—is discussed separately in the next section.

The first difficulty arises from the Bayesian definition of confirmation in terms of probability: a piece of evidence is positively relevant to a theory if and only if it raises the theory's probability. Suppose that we can separate the logical consequences of any theory into two classes: those that contain at least one theoretical term and the so-called observational statements that contain no theoretical terms. The implication rule entails that no theory can be more probable than its consequences. Why, then, would the Bayesian bother with theories at all, if observational statements always have a higher probability than any theory from which they can be derived? One standard answer is that, in addition to empirical warrant, we also value the explanatory power that only theories can provide. But even so, the Bayesian must divorce explanatory power from degree of confirmation. As far as the Bayesian is concerned, observational consequences alone can raise a theory's probability. Explanatory success can do nothing to increase the degree of warranted belief that we should have in a theory, and the probability of a theory can never exceed that of its observational consequences. What Glymour is saying, in effect, is that it is difficult for a

Bayesian to be a scientific realist. Realists typically take the explanatory power of a theory as an important reason for thinking that the theory is true or highly probable. In this way, theories can become more credible than their observational consequences and thus achieve a level of warrant that Bayesians cannot account for.

The second difficulty afflicts both Bayesianism and the H-D model of confirmation. It is the problem of explaining why some pieces of evidence can confirm some parts of a theory but not others, even though the evidence in question is one of the logical consequences of the theory. Glymour regards this problem as crucial to an adequate theory of confirmation, and its solution forms the heart of his own bootstrap model of evidential relevance. The problem can be illustrated by considering the theory consisting of Kepler's three laws of planetary motion. Observations of a single planet can confirm Kepler's first and second laws, but not Kepler's third law. Kepler's third law states that all satellites of the same body (whether they be planets of the sun or moons of Jupiter) obey the equation:

$$T^2 = kr^3,$$

where T is the periodic time of the satellite, r is the mean distance of the satellite from its parent body, and k is a constant for each system of satellites. The fact that at least two satellites are needed to verify Kepler's third law for a system of bodies created difficulties for Isaac Newton in his *Principia*. Using his system of mechanics, Newton was able to prove that any satellite that obeyed Kepler's first and third laws must be subject to an inverse-square-law force. In order to show that this inverse-square-law force was identical to the force we call *gravity* at the surface of the earth, Newton had to prove that it is gravity, attenuated by the inverse square of the earth-moon distance, that is holding the earth's moon in its orbit. But since the earth has only a single satellite, Newton could not appeal to Kepler's third law to prove that the force acting on the moon varies with the inverse square of its distance from the earth. Instead, he had to give a complicated argument involving the precession of the moon's perihelion. It is true that the earth's moon obeys Kepler's third law, but no observation of this single terrestrial satellite can confirm the law. Therein lies the problem for the Bayesian and the hypothetico-deductivist.

Glymour discusses only one Bayesian response to this problem, that by Paul Horwich. Horwich argues that the Bayesian can always engineer the right answer to the question, Which pieces of evidence are relevant to which parts of a theory? by appropriate manipulation of our degrees of belief. But Glymour complains, quite reasonably, that this "solution" is merely a restatement of the problem, since it yields no insight into why our degrees of belief should be distributed in the required manner. As with many of Glymour's other attacks on Bayesianism, the charge is not

that Bayesianism can be refuted by being shown to be inconsistent with some accepted methodological truth but rather that when Bayesians produce the desired consistency it is superficial and ad hoc.

THE PROBLEM OF OLD EVIDENCE

Philosophers of science interested in Bayesianism have probably spilled more ink over the problem of old evidence in the last fifteen years than over any other of the problems discussed by Glymour. The problem arises as follows. On the one hand, scientists often appeal to facts that are already known in support of new theories. Glymour gives as examples Newton's use of Kepler's laws and Einstein's ability to derive the anomalous advance of the perihelion of Mercury from his general relativity theory. Indeed, there is a sense in which any piece of evidence, even evidence that has been newly discovered, must count as "old" if it is being used to confirm a theory, since, unless it is accepted as a well-established fact, it would be useless for the purposes of confirmation. On the other hand, if evidence is "old" in the sense of "known" or, at least "accepted as true," then it would seem that its probability must be 1. Thus, for old evidence, $P(E) = 1$, and regardless of the theory involved, the likelihood $P(E/T)$ must also be 1. But when these values are plugged into Bayes's equation, it follows immediately that E cannot confirm T , because the posterior probability of T given E will be equal to the prior probability of T . Thus, although scientific tradition says that old evidence can, in principle, confirm a theory, Bayesians seem forced to deny this.

Bayesians can make a number of responses to this problem. One reply would be to insist that the scientific tradition is mistaken in thinking that old evidence has any power to confirm a theory, especially when the theory in question was deliberately designed to fit that evidence or explain it (as was the case with the examples of Newton and Einstein cited by Glymour). This response is discussed and criticized in the commentary on chapter 4, in the section "Novel Prediction," and in the sections on Snyder and Achinstein.

If the scientific tradition is not mistaken, the Bayesian has two options: either deny that $P(E) = 1$ or adopt a new criterion for confirmation to replace the standard relevance criterion. Glymour explores versions of both strategies, criticizing the first and cautiously endorsing the second.

The core idea behind the first strategy is that the subjective probability of E should be reckoned, not with respect to one's actual background knowledge, B (which includes E), but with respect to what one's background knowledge would be if E were not yet known. The problem, which Glymour judges to be unsolved, is to give an adequate account of how this counterfactual background knowledge should be defined and how $P(E)$ should be calculated in terms of it. There are two ways to go: either one tries to imagine what one's present background knowledge would have

been if, contrary to fact, E were not known, or one considers what one's background knowledge was in the past, before one learned that E is true. Let us call these the *present proposal* and the *historical proposal*.

With regard to the present proposal, Glymour notes that it is not sufficient merely to say that we should imagine "deleting" E and whatever entails it from B . We need to know, in some detail, how this process is to be carried out, especially when E has been thoroughly integrated into B . Despite their enthusiasm for the first strategy, Colin Howson and Peter Urbach (in their Bayesian response to Glymour) concede that there is no satisfactory formal account of how to delete E from B .²⁵ Part of the problem is that the content of $(B - E)$ varies depending on how the content of B is represented or axiomatized. But they do not see this as an insurmountable barrier, at least when E is an independent item in B , as it will be when E has just been acquired. The Howson-Urbach proposal is that one should take B to be represented in whatever way a particular individual represents her knowledge at a particular time. But they acknowledge that, when E is not an independent item in B , the counterfactual probability of E (and also the probability of E given T) might not be definable and hence will not exist.

Even if the present proposal can be made to yield an acceptable characterization of $(B - E)$, how to calculate the probability of E in terms of $(B - E)$ is still a problem. The usual suggestion is that we should use the total probability rule to express $P(E)$ as a sum of factors of the form $P(T_i) \times P(E/T_i)$. The problem, as Salmon explains in his discussion of the expectedness of evidence, lies in the intractability of the last term in this sum, which involves the catch-all hypothesis, T_k . And as Glymour demonstrates, if one simply deletes the term involving the catch-all hypothesis from the calculation, one's prior and conditional probabilities will inevitably violate the probability axioms and generate absurdities. Thus, calculating the probability of E given the catch-all hypothesis would seem to be impossible, and ignoring the catch-all leads to incoherence.

The historical proposal also has its difficulties. Glymour notes that even if we knew which historical period to choose as the right one, we cannot simply adopt as the value of $P(E)$ whatever degree of belief in E we would have had at that time and retain all our other actual present-day degrees of belief, for that would violate the coherence requirement. Glymour offers a rather complicated solution to this problem of ensuring coherence by making use of Jeffrey conditionalization—a generalization of Bayesian conditionalization proposed by Richard Jeffrey. But this technical difficulty aside, there are serious philosophical objections to the historical proposal. How are we to decide which historical period is the right one? How far back should we go? And when we have made that decision, does it really make sense to suppose that there is a degree of belief that each person would have had in E at that time? Glymour's doubts about the answers to these questions make him as skeptical of the historical

proposal as he is of the present proposal. Thus Glymour rejects the first strategy—the strategy of denying that $P(E) = 1$ —as a solution to the problem of old evidence.

The second strategy concedes that $P(E) = 1$, but offers a new criterion of confirmation to replace the relevance criterion. The second strategy takes its cue from the standard Bayesian assumption (via the coherence condition) that rational agents are logically omniscient. Bayesians assume, quite unrealistically, that rational agents know what all the logical consequences of their beliefs are and assign degrees of belief to them in accordance with the probability axioms. But none of us is even close to being that knowledgeable. In many of the cases involving old evidence, existing theories have been unable to explain some phenomenon (such as the anomalous behavior of Mercury's orbit). A new theory T (such as Einstein's) comes along and we then discover that T entails and is thus able to explain this phenomenon. As a result the new theory is confirmed. The important point is that, while the evidence E is "old," having been known for many years, we have just learned something new, namely, that T entails E , and it is this new piece of information, not E by itself, that confirms T .²⁶ Thus, we can summarize the second strategy by saying that the relevance criterion for confirmation, namely,

E confirms T if and only if $P(T/E \ \& \ B) > P(T/B)$,

should be replaced, at least for the case of old evidence, by the new criterion:

E confirms T if and only if $P(T/E \ \& \ B \ \& \ (T \rightarrow E)) > P(T/E \ \& \ B)$.

Expressed more simply, the new criterion says that:

E confirms T if and only if $P(T/T \rightarrow E) > P(T)$.

In his discussion of the new criterion, Glymour says that the old evidence confirms T in a "vicarious way" (604). But it is misleading to say that, according to the new criterion, E confirms T , however vicariously, when the real confirmational work is being done, not by E , but by the discovery that T entails E . Some authors reject the new criterion because it evades the original problem of whether E literally confirms T . John Earman, for example, objects that the question of whether the astronomical data on Mercury's orbit (the perihelion advance) confirms Einstein's theory has not been answered; it has been replaced by the different question of whether Einstein's confidence in his theory was increased upon his learning that T entails E .²⁷ Moreover, as Earman points out, when most of us learn about Einstein's theory, among the first things we are told, even before we study the theory in detail, is that it entails the peri-

helion advance. So when we later derive this phenomenon from Einstein's theory, we already know that the entailment holds. Does that mean that, for us, the perihelion advance has ceased to be confirming evidence for the general theory of relativity? Other critics object that the new criterion cannot account for the fairly numerous cases in which a theory T was deliberately constructed so as to yield E as one of its consequences. For, in such cases, it can hardly be a discovery that T entails E when that is precisely what the theory was designed to do.²⁸ As far as Glymour is concerned, the new criterion is on the right track, but, like much else in the Bayesian approach, it is superficial. What matters for confirmation, Glymour argues, is the structural relation between T and E . To that extent, the Bayesian is looking in the right place for a solution to the problem of old evidence. But to think that entailment all by itself is an adequate account of that relation is to fall into the same error as the H-D model.

5.4 | Horwich's Defense of Therapeutic Bayesianism

Paul Horwich thinks that many of the criticisms of Bayesianism by Glymour and others are unfair and misdirected, especially those hinging on the near impossibility of assigning numerically precise degrees of belief to scientists and the difficulty of giving rigorous Bayesian explications of every facet of scientific methodology. Bayesianism, Horwich insists, is not intended to be a complete, true, and detailed theory of scientific method. Rather, it should be viewed as a philosophical theory, a contribution to epistemology in general. Horwich follows Wittgenstein (that is, the later Wittgenstein of the *Philosophical Investigations*) in regarding philosophy as therapeutic: its goal is to dissolve problems and resolve paradoxes by untying the knots in our thinking. A good philosophical theory is one that unravels and corrects the misleading assumptions that have led us into confusion and error. The philosophical theories that can achieve this do not have to be especially complex and sophisticated; usually, a simple, idealized model will suffice. And so it is with therapeutic Bayesianism, according to Horwich. By replacing the traditional assumption that belief is an all-or-nothing affair with the idea that there are degrees of belief that conform to the axioms of the probability calculus, we can solve a number of important problems in the philosophy of science. As illustrations of the power of therapeutic Bayesianism, Horwich offers a resolution of the raven paradox and a rationale for our preference for diversity in evidence.

THE RAVEN PARADOX

The raven paradox was made a focus of philosophical inquiry by Carl Hempel in his seminal papers on confirmation theory, where he intro-

duced “All ravens are black” as the paradigm of a scientific hypothesis, thus giving the raven paradox its name.²⁹ Although Hempel drew attention to several confirmational paradoxes, we shall follow Horwich in concentrating solely on one of them, namely that the observation of a nonblack nonraven confirms the hypothesis “All ravens are black.” Somewhat ironically, Hempel, like Horwich, does not regard this result as genuinely paradoxical, since he argues that, when correctly interpreted, observations of white shoes and green chairs (nonblack nonravens) do indeed confirm that all ravens are black. Let us first lay out the three assumptions that generate the alleged paradox and then examine Horwich’s Bayesian resolution of it. For convenience, we shall use the letter *H* to stand for the hypothesis that all ravens are black.

The first assumption is that generalizations of the form “All *R*s are *B*” are confirmed by their positive instances, that is, by anything that satisfies both the antecedent, *R*, and the consequent, *B*. The second assumption is that “All *R*s are *B*” is logically equivalent to “All non-*B*s are non-*R*s.” The third assumption is that if an observation report confirms hypothesis *J*, and *J* is logically equivalent to *K*, then the observation report also confirms *K*. Putting these three assumptions together yields the result that the observation of a nonblack nonraven (such as a white shoe) confirms *H*, that all ravens are black. As Nelson Goodman has remarked, this would seem to open up wonderful prospects for indoor ornithology.³⁰

One way to avoid the paradox is by rejecting one or more of the assumptions generating it. In fact Horwich, like many other Bayesians, rejects the first assumption, the positive-instance principle. The positive-instance principle entails that any observation report of the form “This is a black raven” necessarily confirms the hypothesis, *H*, that all ravens are black. But, Horwich objects, whether or not the observation of a black raven confirms *H* depends, at least in part, on the manner in which the evidence is collected. If, for example, we first pick a black object and then examine it to see whether it is a raven, then the discovery that the object is both black and a raven would do nothing to confirm *H* because the method employed to gather the evidence is incapable of generating a report that could falsify the hypothesis.³¹ Thus Horwich replaces the first assumption with a slightly different principle: “All *R*s are *B*” is confirmed by the positive instance (*Ra* & *Ba*) only if the method for obtaining the report involves first selecting an object that is an *R* and then examining it to see whether it is a *B*. Horwich symbolizes this as (*R** *B*), using an asterisk to indicate the order in which the information was obtained.

Having revised the first assumption, Horwich accepts the other two assumptions and the conclusion of the raven paradox argument. Horwich argues that the appearance of paradox is an illusion that evaporates once we embrace the Bayesian doctrine of degrees of belief and the relevance criterion of confirmation. He notes that the conclusion of the paradox does not say how strongly the observation of a white shoe confirms *H*; it

merely asserts that H is confirmed. The virtue of Bayesianism (as opposed to the H-D model or Glymour's bootstrap model of confirmation) is that it not only permits degrees of confirmation but also allows us to estimate their magnitude. Thus the way is open for the Bayesian to argue that the observation of a white shoe does confirm H but only to a very small degree, a degree that is much smaller than the confirmation conferred on H by the observation of a black raven. In this way, we have a psychological explanation of why many people regard the raven paradox as a genuine paradox when they first encounter it, for most of us are unable to distinguish a very low degree of confirmation from no confirmation at all.

As a measure of the degree to which E confirms H , Horwich adopts the ratio of the posterior probability of H to its prior probability. By Bayes's theorem, this is equal to $P(E/H)$ divided by $P(E)$ —the likelihood of H divided by the expectedness of the evidence.

$$\text{Confirming power of } E = \frac{P(H/E)}{P(H)} = \frac{P(E/H)}{P(E)}$$

As we have noted, Horwich insists that the manner in which evidence is collected is relevant to its confirming power, and he adopts a special notation to reflect this: $(R^* B)$ is the discovery that a randomly selected object that is already known to be a raven is black; $(\sim B^* \sim R)$ is the discovery that a randomly selected object that is already known to be non-black is a nonraven. The order in which the components of each observation are made—a feature that Horwich deems crucial to his analysis—is indicated by the asterisk. To repeat, the observation reports are not simply of the form $(Ra \ \& \ Ba)$ but include information about the method used to generate the report. That said, the likelihoods must both be equal to 1, for if all ravens are black, then the probability that a raven will turn out to be black is 1; similarly, given H , the probability that a nonblack thing will turn out to a nonraven is also 1. So the comparison of the confirming power of $(R^* B)$ with the confirming power of $(\sim B^* \sim R)$ reduces to the inverse ratio of their probabilities: it is the expectedness of evidence $(\sim B^* \sim R)$ divided by the expectedness of evidence $(R^* B)$.

$$\frac{\text{Confirming power of } (R^* B)}{\text{Confirming power of } (\sim B^* \sim R)} = \frac{P(\sim B^* \sim R)}{P(R^* B)}$$

Horwich argues that, given some plausible assumptions about the relative numbers of ravens and black things in the world, our degrees of belief in $(R^* B)$ and $(\sim B^* \sim R)$ prior to our investigation of H can easily be estimated. We assume that, in the world as a whole, the number of ravens is pretty small whereas the number of nonblack things is very large:

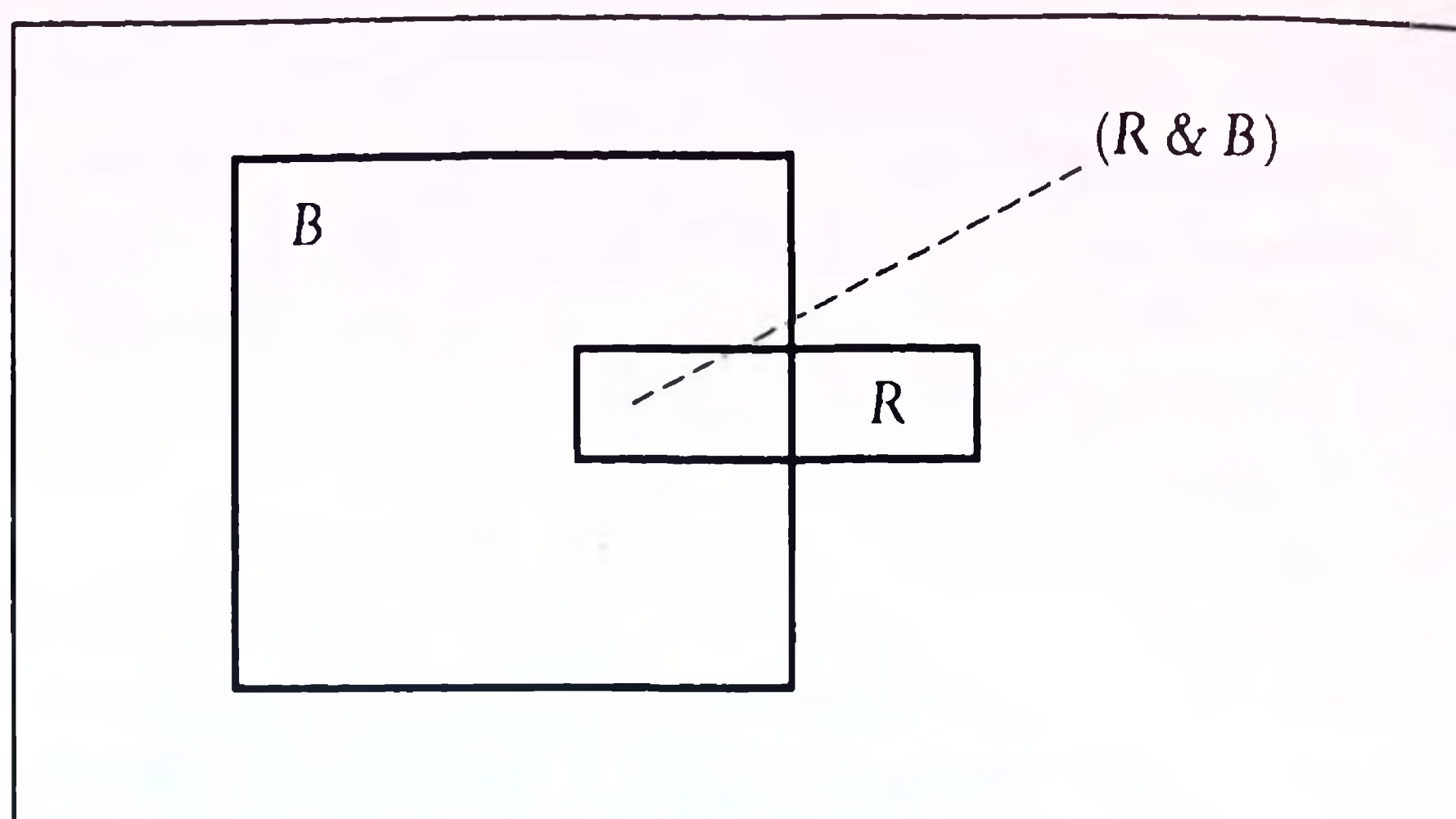


Figure 3

ravens are rare, but nonblack things (and black things, too, for that matter) are abundant. So, Horwich argues, if we randomly select something that is nonblack (which is a very large class of things) the chances that the thing will also be a nonraven must be high (since ravens are scarce and nonravens abundant), and hence our prior degree of belief in $(\sim B^* \sim R)$ must be high. Similarly, he argues, if we randomly select something that is a raven, then the chances that the thing will also be black must be low (given the abundance of nonblack things) and hence our prior degree of belief in $(R^* B)$ must be low. Thus, given the comparison of confirming powers noted above (as the inverse ratio of these probabilities), Horwich concludes that the ratio of $P(\sim B^* \sim R)$ to $P(R^* B)$ is high and so $(R^* B)$ confirms H to a much higher degree than does $(\sim B^* \sim R)$.

Horwich's argument for his Bayesian solution to the raven paradox is not entirely satisfactory. The problem with it can best be grasped by looking at figure 3, which depicts the background assumption that ravens are scarce and nonblack things abundant.

The R -rectangle represents the set of ravens; the area outside of the B -square represents the set of nonblack things. Since the non- B area is so much greater than the R area, Horwich is correct in saying that $P(\sim B^* \sim R)$ must be high. But that same background assumption, by itself, does not warrant the assertion that $P(R^* B)$ must be low. For $P(R^* B)$ depends, not on the size of the R and B areas, but on the fraction of the R -rectangle that overlaps with the B -square. In other words, we need to make some additional assumption about the fraction of ravens that are black. Remarkably, when we add that assumption and analyze the problem more carefully, it turns out that we can still vindicate Horwich's conclusion that black ravens confirm H more strongly than do white shoes. Here is the analysis.

Let x be the fraction of things in the universe that are ravens, let y be the fraction of things that are black, and let α be the fraction of things

that, initially, are believed to be black ravens. In figure 3, the area of the R -rectangle is proportional to x , the area of the B -square is proportional to y , and the area of the intersection between R and B is proportional to α . (If assuming that we have precise beliefs about these numbers seems unrealistic, not to worry—in the final analysis nothing depends on their exact values.) Thus,

$$P(R^*B) = \frac{\alpha}{x}, \text{ and}$$

$$P(\sim B^* \sim R) = \frac{(1 - y) - (x - \alpha)}{(1 - y)}.$$

It follows from these expressions that a necessary and sufficient condition for either (R^*B) or $(\sim B^* \sim R)$ to confirm H is that $x > \alpha$, that there are more ravens than black ravens. Thus, we have proven that $(\sim B^* \sim R)$ confirms H without having to rely on any assumption about the relative abundance of ravens and nonblack things. The only condition for $(\sim B^* \sim R)$ to confirm H is that, prior to the observation, we are not already convinced that all ravens are black. Moreover, we can compare the confirming power of the two reports.

$$\frac{\text{Confirming power of } (R^*B)}{\text{Confirming power of } (\sim B^* \sim R)} = \frac{P(\sim B^* \sim R)}{P(R^*B)} = \left(\frac{x}{\alpha} \right) \frac{(1 - y) - (x - \alpha)}{(1 - y)}.$$

Simple algebra shows that two conditions are necessary and sufficient for the expression on the right-hand side to be greater than 1: the first, as before, is that x is greater than α ; the second is that $(1 - y)$ is greater than x . So Horwich's judgment is vindicated: if we do not already believe that all ravens are black, (R^*B) must support H more strongly than does $(\sim B^* \sim R)$ as long as we also believe that nonblack things are more abundant than ravens. Both sorts of observation must increase our confidence in H , but finding black ravens provides more powerful support.³²

DIVERSITY OF EVIDENCE

Horwich's other illustration of the power of therapeutic Bayesianism is its explanation of the relation between diversity of evidence and strength of support. Why does diverse evidence, E_D , confirm a hypothesis more strongly than narrow evidence, E_N ? In other words, why is it that, for any hypothesis H , $P(H/E_D) > P(H/E_N)$? One might attempt to vindicate the diversity principle by appealing to eliminative induction: as the set of data widens, more rival hypotheses are eliminated, thus enhancing the probability of the survivors. Although he believes that this answer is on the right

track, Horwich thinks that it will not work as it stands because narrow data sets can eliminate just as many hypotheses as can broad ones. The relevant issue, in his view, is not the number but the kind of hypotheses that are excluded. Horwich reasons as follows:

- 1 Diverse data sets tend to eliminate more of the simpler hypotheses than do narrow data sets
- 2 Simpler hypotheses have higher prior probabilities.
- 3 Thus, diverse data sets produce a greater increase in the probability of the surviving hypotheses than do narrow data sets.

In his book *Probability and Evidence*, Horwich gives a semiformal Bayesian treatment of this argument.³³ Let H_S be the straight-line hypothesis in the curve-fitting diagram in Horwich's article (612), and assume that H_S entails both E_D and E_N . Thus

$$P(E_D/H_S) = P(E_N/H_S) = 1.$$

Thus we can write:

$$\frac{P(H_S/E_D)}{P(H_S/E_N)} = \frac{P(E_N)}{P(E_D)} = \frac{P(H_S) + \sum P(E_N/H_i) \times P(H_i)}{P(H_S) + \sum P(E_D/H_i) \times P(H_i)}$$

The numerator and the denominator on the right-hand side of this equation both contain the same sum over all the rival hypotheses to H_S . These rival curves can be divided, roughly, into two classes: the crazy and the gradual (as Horwich labels them in his diagram). But the crazy curves have such low prior probabilities that their contribution to the sum can be ignored. That leaves H_S and the gradual curves, each of which is assumed to have a nonnegligible prior probability because of its relative simplicity. Any gradual curve that goes through E_D is likely also to go through E_N , but there are many gradual curves that go through E_N that go nowhere near E_D . So for most of the gradual curves, $P(E_D/H_i) < P(E_N/H_i)$. Thus, $P(E_N) > P(E_D)$, and so E_D lends greater support to H_S than does E_N .

Despite its ingenuity, Horwich's argument is not unassailable. As we learned from Glymour's criticisms of Bayesianism, proving that simpler hypotheses must be more probable is difficult.³⁴ But Horwich might claim to have shown at least this much: given the widely accepted (but not uncontroversial) assumption that simpler hypotheses are more probable, there is a plausible Bayesian rationale for the diversity principle.³⁵

PROBABILISTIC FOUNDATIONS

In section 2 of his paper, Horwich responds to critics who cast doubt on the foundational assumptions of Bayesianism: the existence of numerical degrees of belief, the coherence condition, the definition of degrees of confirmation in terms of probabilities, and the subjective nature of those probabilities. He acknowledges the difficulties. For example, he concedes that the Dutch book argument for coherence is “far from airtight” (615). Horwich regards the coherence requirement for degrees of belief as a normative idealization, a model of how rational agents ought to constrain their beliefs rather than as a description of how they actually do. Thus, in response to the subjectivity objection, Horwich cites Carnap in support of the contention that, of course, there are objective facts about confirmation even though the probabilities involved are degrees of belief. These facts are objective precisely because confirmation theory, even Bayesian confirmation theory, is normative: it specifies the degree of confidence that a rational agent should place in one proposition given the evidence of another; the right degree of confidence is the one that reflects the true, objective, relation of inductive confirmation between the two propositions.

Horwich also admits that there is room for reasonable disagreement concerning his definition of the degree of confirmation of H as the ratio of $P(H/E)$ to $P(H)$. Other authors, also Bayesians, have advocated different definitions. But Horwich insists that exactly how the degree of confirmation gets defined as a function of probabilities is irrelevant to therapeutic Bayesianism. For to resolve paradoxes and solve most philosophical problems of confirmation, all the therapeutic Bayesian requires is the relevance criterion, shared by all Bayesians, that hypotheses are confirmed if and only if their probability is raised by evidence. In the majority of the cases that interest philosophers, using a specific function to pin an exact number on the degree of confirmation is unnecessary.

MISPLACED SCIENTISM

In the last section of his paper, Horwich broadens his defense of Bayesianism in the philosophy of science by attacking what he calls “misplaced scientism” (618)—the mistake of applying to philosophical theories the standards we use to evaluate scientific theories. Horwich argues that science and philosophy are fundamentally different intellectual activities. Science aims, primarily, at achieving an accurate, true, comprehensive description of the natural world. Philosophy (according to Horwich’s Wittgensteinian conception of it) aims at solving deep puzzles, dissolving paradoxes, and removing the contradictions and confusions that surround concepts such as confirmation, evidence, and simplicity. Given the very different aims of the two endeavors, there is no reason why their theories should be evaluated in the same way. A degree of vagueness, idealization,

informality, and incompleteness that would not be tolerated in the sciences should not disqualify a theory (such as Bayesianism) in the philosophy of science if that theory does a good job of clearing up confusions and providing philosophical insight.

Horwich recognizes that the distinction between science and philosophy is not always sharp, especially in the philosophy of science, which runs the gamut from technical studies in logic and statistical inference to general epistemological theories that are quite informal. Nonetheless, Horwich thinks that we can distinguish two sorts of philosophy of science, theory oriented and problem oriented. Problem-oriented philosophy of science is philosophy properly so-called. It is, to use Horwich's term, "conceptual troubleshooting" (619). To criticize this sort of philosophy of science for being less than fully rigorous or for using highly idealized models is to miss the point. By contrast, theory-oriented philosophy of science is the search for a detailed, accurate, comprehensive account of scientific method. As such, it is really a continuation of science under the guise of naturalized epistemology. Thus, it has the same concern for rigor and completeness that we find in other branches of science. Here, the standards of science are appropriate.

Is the difference between science and philosophy, and between theory-oriented and problem-oriented philosophy of science, great enough to bear the weight that Horwich puts on it? One problem, as he recognizes, is that both sorts of activity have, as one of their goals, the provision of explanations via unifying theories, and often these theories involve simplified models. Idealization for the purpose of explanation is common in both fields. Similarly, in both areas we find the notion that there is some domain of phenomena and problems—whether the field be genetics, biochemistry, or the study of confirmation—of which an adequate theory must give an account. Horwich himself commends Bayesianism for its explanatory successes, for its ability to make sense of a broad range of methodological principles within one unifying framework. But at the same time, he wants us to discount the objections of those who criticize the Bayesian program for the things that it cannot explain by arguing that this is not an appropriate standard to apply to a philosophical theory. This has the appearance of special pleading. If explanatory successes count in favor of a theory, explanatory failures should count against it.

Horwich seems to be on firmer ground when he complains about the unreasonable demand for rigor in areas, such as problem-oriented philosophy of science, where it is either unattainable, inappropriate, or unnecessary. His earlier example of the dispute over the definition of degrees of confirmation is a good example of such misplaced rigor. For most philosophical applications of Bayesianism to puzzles and paradoxes in confirmation theory, no precise confirmation function is needed. Similarly, as far as its philosophical applications are concerned, it is irrelevant that Bayesians cannot measure precise degrees of belief. In science, it would

be unacceptable to build a theory around quantities that cannot be measured or calculated; in problem-oriented philosophy of science it scarcely matters.

5.5 | Summary

In the first section of this commentary, we identified the main elements of the Bayesian approach to confirmation theory. These are, in increasing order of philosophical contentiousness: Bayes's equation, the relevance criterion of confirmation, and the interpretation of the probabilities in that equation as degrees of belief.

Viewed simply as a theorem of the probability calculus, Bayes's equation is entirely uncontroversial. It is a deductive consequence of the axioms of probability, together with a definition of conditional probability. But in order to apply that equation to scientific reasoning, some connection has to be forged between probability and confirmation. The relevance criterion stipulates that a theory, T , is confirmed by evidence, E , whenever E raises the probability of T . Some philosophers of science balk at this. Peter Achinstein, for example, has protested that the relevance criterion entails that whenever Mark Spitz goes swimming, we have confirmed the hypothesis that he will drown, and that this makes a travesty of the notion of confirming evidence. In reply, Bayesians insist on the distinction between a hypothesis being confirmed (which occurs whenever its probability is raised, however slightly) and a hypothesis being acceptable as highly confirmed or true (which requires a posterior probability close to 1). Obviously, no one would accept the hypothesis that Mark Spitz will drown simply because drowning is slightly more likely to occur when he goes swimming; but nonetheless, when Spitz goes swimming, this increases the chances that he will drown and thus, to a small degree, confirms the hypothesis.

Even when one accepts the relevance criterion, there is still the question of what $P(T)$, $P(E/T)$, $P(E)$, and $P(T/E)$ stand for in Bayes's equation. These four quantities are usually referred to as the prior probability of theory T , the likelihood of T , the expectedness of evidence E , and the posterior probability of T . The official Bayesian view is that all of these quantities are subjective degrees of belief, that is, they are the actual degrees of belief of real people as measured, for example, by their willingness to bet at well-defined odds on various propositions. The appearance of subjectivism is mitigated somewhat by the stipulation that the people in question have to be rational and that rationality requires that their degrees of belief satisfy the probability axioms. The requirement of conformity to the probability axioms is called the *condition of coherence*, and Bayesians usually offer a Dutch book argument to justify it. The other constraint

imposed by Bayesians is conditionalization: whatever a person's prior degree of belief in T , when new evidence, E , is acquired, that person should replace the old prior probability of T with the posterior probability of T given E , calculated using Bayes's theorem: posterior probabilities replace old prior probabilities. Again, Bayesians have appealed to Dutch book arguments in defense of this requirement.

Wesley Salmon is attracted to the Bayesian approach because it seems to represent a definite advance over the H-D model. Unlike the H-D model, the Bayesian approach can be applied to statistical theories, and it holds out the promise of being able to solve the Duhem problem (the problem of allocating blame within a group of hypotheses for a failed prediction) and the tacking paradox (otherwise known as the problem of irrelevant conjunction). Salmon also thinks that Bayesianism can be used to reconcile Thomas Kuhn's historically oriented approach to understanding science with logical empiricism. The key to this reconciliation, in Salmon's view, lies in the role of prior probabilities in determining degrees of confirmation. The factors influencing the judgment of scientists about theories—Kuhnian values such as consistency, simplicity, and fruitfulness—can lead to reasonable disagreements about how well a theory is confirmed by a piece of evidence by giving rise to different estimates of the theory's prior probability. Even though they may be using the same Bayesian algorithm, scientists generate different outputs (confirmation judgments) because their inputs (prior probabilities) differ.

Despite his enthusiasm for Bayes's theorem as a tool for understanding confirmation, Salmon rejects the unfettered subjectivism of many Bayesians. At bottom, he thinks, prior probabilities should be guided by experience. They should not reflect mere prejudice and subjective whim. But Salmon takes comfort in the fact that, in the long run, the initial choice of prior probabilities becomes irrelevant. For as evidence accumulates, the values of the posterior probability of a hypothesis converge, thus "swamping out" or "washing out" individual differences among the priors. One important condition for this sort of convergence is that we never assign to any hypothesis the extreme values of 0 or 1. Another condition is that we agree about the values of likelihoods.

In Salmon's view, the biggest difficulty in applying Bayes's theorem to scientific theories lies in the calculation of $P(E)$, the expectedness of the evidence. In fact, Salmon judges this problem to be insoluble. He therefore advocates a modified version of the Bayesian approach, employing an algorithm for theory preference that does not require us to calculate $P(E)$. This places a significant limitation on the scope of Salmon's analysis, for his algorithm permits the comparison of rival theories only when the evidence for each is the same.

Clark Glymour attacks Bayesianism on several fronts. He points out that the foundations of Bayesianism—the appeal to betting behavior, the Dutch book arguments, and the convergence theorem—promise more

than they can deliver and are less than fully secure. Many Bayesians admit as much but contend that the real strength of Bayesianism lies in its ability to explicate methodological principles such as our preference for relevant evidence and simple theories. Glymour doubts this. He examines several different Bayesian analyses of simplicity and finds each of them inadequate. With regard to our preference for evidence that is relevant to a hypothesis (and not merely a logical consequence of it), Glymour concedes that Bayesians can make their theory fit our intuitions by appropriately assigning values to prior probabilities. But, he complains, this kind of gerrymandering yields no insight into why some pieces of evidence (and not others) are confirmationally relevant to some parts of a theory (and not to others). At best, the Bayesian account of these matters is superficial.

A particularly difficult problem for the Bayesian account is the problem of old evidence. Scientific theories are often regarded as confirmed when they can explain some phenomenon or law that is already well known and accepted. But if the evidence, E , is old in this sense, then it would seem that its probability should be 1. Unfortunately, making $P(E) = 1$ in Bayes's equation would prevent E from affecting the probability of any theory, T , that explains it. Thus, old evidence should not be able to confirm T .

Not everyone agrees that assigning old evidence a probability of 1 and thus denying it any power to confirm theories is a mistake. (See, for example, the discussion of predictionism in the sections of the commentary on chapter 4 devoted to Hempel, Snyder, and Achinstein.) But most Bayesians have tried to solve the problem by arguing either that old evidence should not be given a probability of 1 simply because it is already known or that we should adopt a new criterion of confirmation. Glymour discusses two versions of the first strategy, which we have labeled the present proposal and the historical proposal. Glymour criticizes both. The present proposal advocates that we calculate $P(E)$ by trying to imagine what probability we would assign to E were E (and whatever entails it) to be deleted from our background knowledge. As Glymour explains, this operation of "deletion" is not easy to perform. It can also lead to violations of the probability axioms if we try to calculate $P(E)$ using the total probability rule to sum over all the hypotheses that can explain E . The historical proposal avoids this problem by taking as the value of $P(E)$ whatever the degree of belief in E was (or would have been) before it was discovered. This proposal, too, runs into difficulties, such as the problem of choosing the "right" historical period.

The second strategy, to which Glymour himself inclines, takes a different tack. It suggests that the old evidence itself is not what confirms the theory. Rather, it is the discovery that the theory can explain the evidence (by entailing it) that provides the confirmation. Since scientists are not logically omniscient, they can genuinely discover that some theory T entails evidence E . Despite its ingenuity, Glymour criticizes this proposal,

like much else in the Bayesian system, for not getting to the epistemological heart of the matter. As he emphasizes, there is more to the confirmational relevance of E to T than the mere logical derivability of E from T .

Paul Horwich undertakes to defend Bayesianism against critics such as Glymour by arguing for its value as a philosophical theory that can solve conceptual problems. Horwich gives two illustrations of the power of what he calls “therapeutic Bayesianism”: the resolution of the raven paradox and the explanation of why it is desirable to seek diversity in evidence.

The raven paradox follows from seemingly plausible principles of confirmation. Apparently, the observation of a nonblack thing that is not a raven should confirm the hypothesis that all ravens are black. Carl Hempel, who first drew attention to this result, regarded it as correct: the observation of a white shoe does confirm that all ravens are black, but only to a very small degree. Because the degree of confirmation is so very small, we have a psychological tendency to mistake it for no confirmation at all. Horwich attempts to vindicate Hempel’s position by giving a simple Bayesian analysis incorporating assumptions about the relative size of the class of ravens as compared with the (very much larger) class of things that are not black.

Horwich uses a curve-fitting diagram to illustrate his Bayesian explanation of why diverse evidence confirms a hypothesis more strongly than does narrow evidence. The gist of his argument is that diverse evidence confirms more effectively because it eliminates a greater number of the simple hypotheses. A crucial premise in this argument is that simple hypotheses have higher prior probabilities. As we learned in chapters 4 and 5, it is very difficult to connect simplicity with truth or the probability of truth. But if Horwich’s Bayesian argument is to have any power to convince non-Bayesians, the prior probabilities to which it appeals must be more than mere expressions of subjective preference.

In the remaining two sections of his paper, Horwich responds to some criticisms of the foundations of Bayesianism and concludes by attacking what he calls “misplaced scientism” (618)—the error of imposing on a philosophical theory the same standards of quantitative rigor that we demand of theories in science. Horwich recognizes these standards—reasonable in science and even within some branches of the philosophy of science, where the aim is to give an accurate, comprehensive description of some domain of phenomena. But when we are dealing with a philosophical theory, the primary function of which is to solve conceptual problems, Horwich regards these standards as inappropriate. Thus, he rejects as unwarranted some of the complaints made against Bayesianism, such as the difficulty of measuring precise degrees of belief or specifying a well-defined confirmation function. For many of the applications of the Bay-

esian theory in philosophy, Horwich judges this lack of completeness and rigor to be irrelevant.

■ | Notes

1. An argument for preferring the difference analysis of degree of confirmation is given in footnote 9, below.
2. A small complication: since probabilities cannot exceed one, the area of each circle must be divided by the area of the rectangle that encloses all the circles. When there are just two circles, A and B , as in figure 2, the enclosing rectangle represents the proposition $(A \vee B \vee \sim A \vee \sim B)$. Since this is a necessary truth, its probability is one and so, too, is its area, when expressed in suitable units.
3. Here is a simple illustration. There are one hundred passengers on a cruise ship of whom forty speak French (and no English) and ten speak Swedish (and no English); all the rest speak only English. Exactly two passengers, Ingmar and Chantal, speak both French and Swedish. What is the probability that a randomly selected passenger speaks either French or Swedish? If we simply add $P(F)$ and $P(S)$ to get 0.5, then we will have counted Ingmar and Chantal twice, since they belong to both groups. Therefore, we must subtract 0.02 to get the correct answer of 0.48. Alternatively, we could calculate from the information given that the number of passengers who speak English is fifty-two (not fifty!) and hence that forty-eight out of a hundred is the fraction of passengers who speak either French or Swedish.
4. In some treatments of probability, the definition of $P(A/B)$ is regarded as a fourth axiom. See, for example, footnote 1 of Horwich's article.
5. Thus when someone refers to $P(E/H)$ as the likelihood of evidence E on hypothesis H , he is using the term *likelihood* informally, as a synonym for *probability*. Strictly speaking, $P(E/H)$ should be referred to as the likelihood of H , or the likelihood of H on E .
6. We leave to one side Richard Jeffrey's generalization of the Bayesian rule of conditionalization (discussed in Glymour's article), which handles cases in which the evidence, E , is not 100 percent certain given the observations on which it is based.
7. Paul Teller, "Conditionalization and Observation," *Synthese* 26 (1973): 218–58. Teller attributes the diachronic Dutch book argument to David Lewis.
8. Wesley C. Salmon, "The Appraisal of Theories: Kuhn Meets Bayes," in *PSA* 1990, vol. 2 (East Lansing, Mich.: Philosophy of Science Association), 325. Evidently, Salmon did not read very far, since Kuhn first casts doubt on the traditional distinction between the two contexts on pages 8–9 of his book.
9. According to the other version, the degree of confirmation depends on the ratio of the two probabilities. But for deterministic theories, this version entails that theories such as T and $(T \& I)$ are confirmed to the same degree by E , since in both cases, $P(T/E \& B)$ divided by $P(T/B)$, and $P(T \& I/E \& B)$ divided by

$P(T \& I/B)$, is simply the reciprocal of $P(E/B)$. This is a compelling reason for Bayesians to prefer the difference analysis of degree of confirmation.

10. See Jon Dorling, "Bayesian Personalism, the Methodology of Scientific Research Programmes, and Duhem's Problem," *Studies in History and Philosophy of Science* 10 (1979): 177–87.

11. Wesley C. Salmon, *The Foundations of Scientific Inference* (Pittsburgh, Pa.: University of Pittsburgh Press, 1967).

12. For these and other criticisms of the attempt by Reichenbach and Salmon to interpret prior probabilities as frequencies, see Mary B. Hesse, *The Structure of Scientific Inference* (Berkeley, Calif.: University of California Press, 1974), 106, and Alvin Plantinga, "The Probabilistic Argument from Evil," *Philosophical Studies* 35 (1979): 1–53.

13. Indeed, if the degree of confirmation is proportional to the difference between prior and posterior probabilities for each theory given evidence E , then in cases where T_1 and T_2 are statistical theories neither of which logically entails E , $P(E/B)$ will not cancel out when we divide the difference between $P(T_1/E \& B)$ and $P(T_1/B)$ by the difference between $P(T_2/E \& B)$ and $P(T_2/B)$. For statistical theories, the ratio of the degrees of confirmation is equal to Salmon's ratio of the posterior probabilities only when their likelihoods are the same; that is, only when $P(E/T_1 \& B)$ equals $P(E/T_2 \& B)$.

14. Just to emphasize that the comparative support measured by Salmon's algorithm is not a comparison of degrees of inductive confirmation, we point out that E might disconfirm both T_1 and T_2 , and yet, by Salmon's algorithm, T_2 could be judged preferable to T_1 , because the posterior probability of T_2 given E is greater than the posterior probability of T_1 given E .

15. For these and other criticisms of Dutch book arguments, see Henry E. Kyburg, Jr., "Subjective Probability: Criticisms, Reflections, and Problems," *Journal of Philosophical Logic* 7 (1978): 157–80.

16. If there is only one black ball in the urn, the chances of getting six black balls in ten draws (with replacement) is $\frac{10!}{6!4!} \left(\frac{1}{80}\right)^6 \left(\frac{79}{80}\right)^4$.

17. Although it is not included in his extant writings, the principle that "entities should not to be multiplied beyond necessity" is traditionally attributed to the scholastic philosopher William of Ockham (c. 1285–1347). *Ockham* is also spelt *Occam*, and the principle is universally referred to as *Ockham's* (or *Occam's*) razor.

18. C. Glymour, "The Epistemology of Geometry," *Nous* 11 (1977): 227–51.

19. Glymour also regards it as a telling strike against hypothetico-deductivism that, without some significant embellishment, it lacks the resources to rule out deoccamized alternatives to theories that are scientifically legitimate.

20. Glymour's argument is not entirely convincing on this last point, since the Bayesian could admit the relativization of deoccamization to classes of evidence, yet still assign low prior probabilities to those theories that qualify as deoccamized relative to a specified class of evidence and higher prior probabilities to those that do not. And even if it should turn out that most respectable scientific theories

have some degree of theoretical indeterminacy with respect to most bodies of evidence, it might be possible to specify degrees of deoccamization and assign the highest prior probabilities to theories with the lowest degrees.

21. See Malcom Forster, "Bayes and Bust: Simplicity as a Problem for a Probabilist's Approach to Confirmation," *British Journal for the Philosophy of Science* 46 (1995): 399–424.

22. See Mary B. Hesse, *The Structure of Scientific Inference*, 153–54, 229–39. Hesse distinguishes two senses of simplicity: content and economy. The focus in Glymour's article and this commentary is on just one sort of economy, namely, economy of parameters and properties. See Hesse's book for her treatment of other sorts of economy (such as economy of mathematical form, theoretical premises, and ontology).

23. For Rosenkrantz's definition of average likelihood and a simple illustration, see pp. 74–76 of Roger D. Rosenkrantz, "Why Glymour Is a Bayesian," in *Testing Scientific Theories*, ed. John Earman, vol. 10, *Minnesota Studies in the Philosophy of Science* (Minneapolis: University of Minnesota Press, 1983), 69–97.

24. For Rosenkrantz's reply, see Roger D. Rosenkrantz, "Why Glymour Is a Bayesian." In response to Glymour's charge that he has failed to show "that in curve fitting the average likelihood of a linear hypothesis is greater than the average likelihood of a quadratic or higher-degree hypothesis," Rosenkrantz writes. "But of course I don't want to show *that*, for it isn't true! What can be shown is that the average likelihood of the quadratic family will be higher than that of the linear family when the data fit the quadratic hypothesis *sufficiently better* than the linear one, whereas the latter will enjoy higher average likelihood when the two families fit equally well" (82).

25. Colin Howson and Peter Urbach, *Scientific Reasoning: The Bayesian Approach*, 2d ed. (La Salle, Ill.: Open Court, 1993), 405–7.

26. Since the publication of Glymour's book, a number of authors have defended versions of the second strategy. See Daniel Garber, "Old Evidence and Logical Omniscience in Bayesian Confirmation Theory," and Richard Jeffrey, "Bayesianism with a Human Face," in *Testing Scientific Theories*, ed. John Earman, vol. 10, *Minnesota Studies in the Philosophy of Science* (Minneapolis: University of Minnesota Press, 1983), 99–131, 133–56; Ilkka Niiniluoto, "Novel Facts and Bayesianism," *British Journal for the Philosophy of Science* 34 (1983): 375–79. See also chapter 5 of John Earman, *Bayes or Bust?*

27. John Earman, *Bayes or Bust?* 130–31.

28. Colin Howson and Peter Urbach, *Scientific Reasoning: The Bayesian Approach*, 407–8.

29. See especially, Carl G. Hempel, "Studies in the Logic of Confirmation," *Mind* 54 (1945): 1–26, 97–121; reprinted in *Aspects of Scientific Explanation* (New York: Macmillan, 1965), 3–46.

30. Nelson Goodman, *Fact, Fiction, and Forecast*, 4th ed. (Cambridge, Mass.: Harvard University Press, 1983), 70–71.

31. For other Bayesian criticisms of the positive instance principle, see I. J. Good, "The White Shoe Is a Red Herring," *British Journal for the Philosophy of Science*

17 (1967): 322; Richard G. Swinburne, "The Paradoxes of Confirmation—A Survey," *American Philosophical Quarterly* 8 (1971): 318–30; Roger D. Rosenkrantz, *Inference, Method and Decision* (Dordrecht, Netherlands: D. Reidel, 1977), ch. 2; Roger D. Rosenkrantz, "Does the Philosophy of Induction Rest on a Mistake?" *Journal of Philosophy* 79 (1982): 78–97. One of Good's counterexamples to the positive instance principle relies on background information. Suppose we know that we are in one of two worlds: in one there are a hundred black crows, no crows that are not black, and a million other birds; in the other there are a thousand black crows, a single white crow, and a million other birds. A bird selected at random turns out to be a black crow. This is strong evidence that we are in the second world and thus the observation of a black crow undermines the hypothesis that all crows are black. (The evidence, E , for W_2 is strong as measured by Salmon's algorithm. For, if W_1 and W_2 have the same prior probability, $P(W_2/E)$ is about ten times greater than $P(W_1/E)$.)

32. For further illustrations of using background assumptions to resolve paradoxes in confirmation theory, see Richard G. Swinburne, "The Paradoxes of Confirmation—A Survey," and his book, *An Introduction to Confirmation Theory* (London: Methuen and Co., 1973).

33. Paul Horwich, *Probability and Evidence* (Cambridge: Cambridge University Press, 1982).

34. As Horwich notes in his article, he defends the simplicity postulate in his book, *Probability and Evidence*.

35. For a criticism of Horwich's argument, see Andrew Wayne, "Bayesianism and Diverse Evidence," *Philosophy of Science* 62 (1995): 111–21. Daniel Steel has replied to this criticism in "Bayesianism and the Value of Diverse Evidence," *Philosophy of Science* 63 (1996): 666–74.



6 | Models of Explanation

INTRODUCTION

It is taken for granted in modern science that one of the main functions of scientific theories is to explain. What gets explained can vary widely from particular events and facts (e.g., why did the space shuttle *Challenger* explode? what caused the extinction of the dinosaurs? why do the equinoxes precess?) to empirical laws and theories (e.g., why do planets obey Kepler's second law? why do liquids maintain a constant temperature as they boil? why do elements from column 1 of the periodic table combine so vigorously with elements from column 7?). Laws and theories can be used to explain events and facts; higher-level laws and theories can be used to explain lower-level laws; and some theories can be used to explain others. In this chapter the main focus is on the explanation of particular events and facts. The explanation of laws and theories is addressed separately in chapter 8, on reduction.

Customarily, the thing that gets explained (or a sentence describing it) is called the *explanandum*. (In Latin *explanandum* means simply "the thing to be explained.") The thing that does the explaining is called the *explanans* (again, a Latin word meaning "that which explains"). There is a simple way to remember these terms: the *explanans* provides the answer to the question—the *um?*—raised in the *explanandum*. Typically, in offering a model of explanation, a philosopher of science will tell us what conditions a group of sentences must satisfy in order to be an *explanans* and how they must be related to the *explanandum* in order to constitute a genuine explanation.

Without doubt, the most influential claim about explanation in the twentieth century has been the covering law thesis of Carl Hempel, according to which genuine explanations are arguments that must contain empirical laws. Hempel's two principal covering law models of explana-

tion, defended and attacked in this chapter's readings, are the deductive-nomological (D-N) model and the inductive-statistical (I-S) model.¹

In "The Value of Laws: Explanation and Prediction," Rudolf Carnap sheds light on the motivation for Hempel's covering law thesis. He relates conversations that took place in the early 1930s among members of the Vienna Circle who were dissatisfied with the *entelechy* theory advocated by the German biologist, Hans Driesch. Carnap and his colleagues diagnosed the main flaw in Driesch's theory to be its failure to provide testable laws, without which Driesch's theory could not really explain anything. Subsequently, Hempel included this requirement as a central condition of his two models of explanation, described in "Two Basic Types of Scientific Explanation."

According to Hempel's D-N model, an important class of scientific explanations are deductively valid arguments, with a statement of at least one law among the premises and a description of the event to be explained in the conclusion. If, instead of explaining an event, we had predicted it, the argument involved would have been very similar. This formal similarity between D-N explanations and predictions leads Hempel to espouse the doctrine discussed in "The Thesis of Structural Identity," that all adequate explanations are potentially predictions and all adequate predictions are potentially explanations. Critics of the D-N model have attacked both parts of this thesis, and Hempel responds to these criticisms—referred to, collectively, as the *symmetry objection*—in his article.

Although Hempel proposed the D-N model in 1948, it was not until 1962 that he considered in detail those scientific explanations that use statistical laws to explain particular events. In "Inductive-Statistical Explanation," Hempel sets out his I-S model of these explanations. Hempel construes them as inductive arguments in which the explanans does not entail the explanandum but establishes it with reasonably high probability. From the outset, Hempel realized that I-S explanations face a problem that does not arise for D-N explanations, namely the problem of ambiguity: several statistical laws might be used to explain some particular event with each of these laws assigning a different probability to the event in question. Particularly disturbing, in Hempel's view, is the possibility that we could have explained the outcome regardless of whether or not the event actually occurred. Hempel's proposed solution to the problem of ambiguity for statistical explanations is his requirement of maximal specificity. But this requirement leads Hempel to argue that, unlike D-N explanations, I-S explanations are, by their very nature, relative to the state of scientific knowledge at a particular time.

There have been many searching criticisms of Hempel's two models of explanation. In "Arguments, Laws, and Explanation," David-Hillel Ruben focuses on an important group of criticisms that go under the heading of the *irrelevance objection*. Ruben charges that Hempel's models are flawed because they can be satisfied by arguments with premises that

are explanatorily irrelevant to their conclusions. Usually, this irrelevance arises because the premises fail to identify the cause of the explanandum.² Ruben discusses ways of trying to meet the irrelevance objection by imposing the requirement—referred to as the causal condition—that the explanans has to mention the cause of the explanandum. He ends up espousing the single-statement view of explanation, which, contrary to Hempel's covering law thesis, denies that explanations are arguments or that they have to involve laws.

Like David-Hillel Ruben, Peter Railton also repudiates several of the main features of Hempel's covering law models. In "A Deductive-Nomological Model of Probabilistic Explanation," Railton rejects Hempel's contention that probabilistic explanations are arguments, along with Hempel's high-probability and maximal specificity requirements. Railton's own model of explanation—the deductive-nomological model of his title—has as its central feature the specification of the mechanism that brings about the event to be explained. The virtues of this account, according to Railton, are that it permits the explanation of improbable events and makes statistical explanations entirely objective.

■ | Notes

1. Hempel claims that his models explicate the concept of explanation not only in the physical sciences and biology, but also in the social sciences and history. The extension of Hempel's account to the explanation of human action and behavior is controversial and falls outside of the scope of the present volume. For a good introduction to the issues involved, see Carl G. Hempel, "The Function of General Laws in History," *Journal of Philosophy* 39 (1942): 35–42; reprinted in Carl G. Hempel, *Aspects of Scientific Explanation* (New York: Free Press, 1965), 231–43; Carl G. Hempel, "Aspects of Scientific Explanation," in *Aspects of Scientific Explanation*, sec. 6–10; William H. Dray, *Laws and Explanation in History* (Oxford: Oxford University Press, 1957); Patrick Gardiner, ed., *Theories of History* (New York: Free Press, 1959); Alan Donagan, "The Popper-Hempel Theory Reconsidered," in *Philosophical Analysis and History*, ed. W. H. Dray (New York: Harper and Row, 1966), 127–59; David Papineau, *For Science in the Social Sciences* (New York: St. Martin's Press, 1978); Merrilee H. Salmon, "Explanation in the Social Sciences," in *Scientific Explanation*, ed. Philip Kitcher and Wesley C. Salmon, vol. 13, *Minnesota Studies in the Philosophy of Science* (Minneapolis: University of Minnesota Press, 1989), 384–409.

2. To avoid circumlocution, we follow Hempel in using the term *explanandum* to refer to either the event-to-be-explained or a sentence describing that event.

The Value of Laws: Explanation and Prediction

In the nineteenth century, certain Germanic physicists, such as Gustav Kirchhoff and Ernst Mach, said that science should not ask "Why?" but "How?" They meant that science should not look for unknown metaphysical agents that are responsible for certain events, but should only describe such events in terms of laws. This prohibition against asking "Why?" must be understood in its historical setting. The background was the German philosophical atmosphere of the time, which was dominated by idealism in the tradition of Fichte, Schelling, and Hegel. These men felt that a description of how the world behaved was not enough. They wanted a fuller understanding, which they believed could be obtained only by finding metaphysical causes that were behind phenomena and not accessible to scientific method. Physicists reacted to this point of view by saying: "Leave us alone with your why-questions. There is no answer beyond that given by the empirical laws." They objected to why-questions because they were usually metaphysical questions.

Today the philosophical atmosphere has changed. In Germany there are a few philosophers still working in the idealist tradition, but in England and the United States it has practically disappeared. As a result, we are no longer worried by why-questions. We do not have to say, "Don't ask why", because now, when someone asks why, we assume that he means it in a scientific, nonmetaphysical sense. He is simply asking us to explain something by placing it in a framework of empirical laws.

When I was young and part of the Vienna Circle, some of my early publications were written as a reaction to the philosophical climate of

FROM Rudolf Carnap, *Philosophical Foundations of Physics*, ed. Martin Gardner (New York: Basic Books, 1966), 12-16. This reading incorporates corrections made by the editor, Martin Gardner, when Carnap's book was reprinted in 1995 by Dover Publications, Inc., under the title, *An Introduction to the Philosophy of Science*.

German idealism. As a consequence, these publications and those by others in the Vienna Circle were filled with prohibitory statements similar to the one I have just discussed. These prohibitions must be understood in reference to the historical situation in which we found ourselves. Today, especially in the United States, we seldom make such prohibitions. The kind of opponents we have here are of a different nature, and the nature of one's opponents often determines the way in which one's views are expressed.

When we say that, for the explanation of a given fact, the use of a scientific law is indispensable, what we wish to exclude especially is the view that metaphysical agents must be found before a fact can be adequately explained. In prescientific ages, this was, of course, the kind of explanation usually given. At one time, the world was thought to be inhabited by spirits or demons who are not directly observable but who *act* to cause the rain to fall, the river to flow, the lightning to flash. In whatever one saw happening, there was something—or, rather, *somebody*—responsible for the event. This is psychologically understandable. If a man does something to me that I do not like, it is natural for me to make him responsible for it and to get angry and hit back at him. If a cloud pours water over me, I cannot hit back at the cloud, but I can find an outlet for my anger if I make the cloud, or some invisible demon behind the cloud, responsible for the rainfall. I can shout curses at this demon, shake my fist at him. My anger is relieved. I feel better. It is easy to understand how members of prescientific societies found psychological satisfaction in imagining agents behind the phenomena of nature.

In time, as we know, societies abandoned their mythologies, but sometimes scientists replace the spirits with agents that are really not much different. The German philosopher Hans Driesch, who died in 1941, wrote many books on the philosophy of science. He was originally a prominent biologist, famed for his work on certain organismic responses, including regeneration in sea urchins. He cut off parts of their bodies and observed in which stages of their growth and under what conditions they were able to grow new parts. His scientific work was important and excellent. But Driesch was also interested in philosophical questions, especially those dealing with the foundations of biology, so eventually he became a professor of philosophy. In philosophy also he did some excellent work, but there was one aspect of his philosophy that I and my friends in the Vienna Circle did not regard so highly. It was his way of *explaining* such biological processes as regeneration and reproduction.

At the time Driesch did his biological work, it was thought that many characteristics of living things could not be found elsewhere. (Today it is seen more clearly that there is a continuum connecting the organic and inorganic worlds.) He wanted to explain these unique organismic features, so he postulated what he called an "entelechy". This term had been introduced by Aristotle, who had his own meaning for it, but we need not

discuss that meaning here. Driesch said, in effect: "The entelechy is a certain specific force that causes living things to behave in the way they do. But you must not think of it as a *physical* force such as gravity or magnetism. Oh, no, nothing like that."

The entelechies of organisms, Driesch maintained, are of various kinds, depending on the organism's stage of evolution. In primitive, single-celled organisms, the entelechy is rather simple. As we go up the evolutionary scale, through plants, lower animals, higher animals, and finally to man, the entelechy becomes more and more complex. This is revealed by the greater degree to which phenomena are integrated in the higher forms of life. What we call the "mind" of a human body is actually nothing more than a portion of the person's entelechy. The entelechy is much more than the mind, or, at least, more than the conscious mind, because it is responsible for everything that every cell in the body does. If I cut my finger, the cells of the finger form new tissue and bring substances to the cut to kill incoming bacteria. These events are not consciously directed by the mind. They occur in the finger of a one-month-old baby, who has never heard of the laws of physiology. All this, Driesch insisted, is due to the organism's entelechy, of which mind is one manifestation. In addition, then, to scientific explanation, Driesch had an elaborate theory of entelechy, which he offered as a *philosophical* explanation of such scientifically unexplained phenomena as the regeneration of parts of sea urchins.

Is this an explanation? I and my friends had some discussions with Driesch about it. I remember one at the International Congress for Philosophy, at Prague, in 1934. Hans Reichenbach and I criticized Driesch's theory, while he and others defended it. In our publications we did not give much space to this criticism because we admired the work Driesch had done in both biology and philosophy. He was quite different from most philosophers in Germany in that he really wanted to develop a scientific philosophy. His entelechy theory, however, seemed to us to lack something.

What it lacked was this: the insight that you cannot give an explanation without also giving a law.

We said to him: "Your entelechy—we do not know what you mean⁴ by it. You say it is not a physical force. What is it then?"

"Well", he would reply (I am paraphrasing his words, of course), "you should not be so narrow-minded. When you ask a physicist for an explanation of why this nail suddenly moves toward that bar of iron, he will tell you that the bar of iron is a magnet and that the nail is drawn to it by the force of magnetism. No one has ever seen magnetism. You see only the movement of a little nail toward a bar of iron."

We agreed. "Yes, you are right. Nobody has seen magnetism."

"You see", he continued, "the physicist introduces forces that no one can observe—forces like magnetism and electricity—in order to explain

certain phenomena. I wish to do the same. Physical forces are not adequate to explain certain organic phenomena, so I introduced something that is force-like but is not a physical force because it does not act the way physical forces act. For instance, it is not spatially located. True, it acts on a physical organism, but it acts in respect to the entire organism, not just to certain parts of it. Therefore, you cannot say where it is located. There is no location. It is not a physical force, but it is just as legitimate for me to introduce it as it is for a physicist to introduce the invisible force of magnetism."

Our answer was that a physicist does not explain the movement of the nail toward the bar simply by introducing the word "magnetism". Of course, if you ask him why the nail moves, he may answer first by saying that it is due to magnetism; but if you press him for a fuller explanation, he will give you laws. The laws may not be expressed in quantitative terms, like the Maxwell equations that describe magnetic fields; they may be simple, qualitative laws with no numbers occurring in them. The physicist may say: "All nails containing iron are attracted to the ends of bars that have been magnetized." He may go on to explain the state of being magnetized by giving other nonquantitative laws. He may tell you that iron ore from the town of Magnesia (you may recall that the word "magnetic" derives from the Greek town of Magnesia, where iron ore of this type was first found) possesses this property. He may explain that iron bars become magnetized if they are stroked a certain way by naturally magnetic ores. He may give you other laws about conditions under which certain substances can become magnetized and laws about phenomena associated with magnetism. He may tell you that if you magnetize a needle and suspend it by the middle so that it swings freely, one end will point north. If you have another magnetic needle, you can bring the two north-pointing ends together and observe that they do not attract but repel each other. He may explain that if you heat a magnetized bar of iron, or hammer it, it will lose magnetic strength. All these are qualitative laws that can be expressed in the logical form, "if . . . then . . ." The point I wish to emphasize here is this: it is not sufficient, for purposes of explanation, simply to introduce a new agent by giving it a new name. You must also give laws.

Driesch did not give laws. He did not specify how the entelechy of an oak tree differs from the entelechy of a goat or giraffe. He did not classify his entelechies. He merely classified organisms and said that each organism had its own entelechy. He did not formulate laws that state under what conditions an entelechy is strengthened or weakened. Of course he described all sorts of organic phenomena and gave general rules for such phenomena. He said that if you cut a limb from a sea urchin in a certain way, the organism will not survive; if you cut it another way, the organism will survive, but only a fragmentary limb will grow back. Cut in still

another way and at a certain stage in the sea urchin's growth, it will regenerate a new and complete limb. These statements are all perfectly respectable zoological laws.

"What do you add to these empirical laws", we asked Driesch, "if after giving them you proceed to tell us that all the phenomena covered by those laws are due to the sea urchin's entelechy?"

We believed that nothing was added. Since the notion of an entelechy does not give us new laws, it does not explain more than the general laws already available. It does not help us in the least in making new predictions. For these reasons we cannot say that our scientific knowledge has increased. The concept of entelechy may at first seem to add something to our explanations; but when we examine it more deeply, we see its emptiness. It is a pseudoexplanation.

It can be argued that the concept of entelechy is not useless if it provides biologists with a new orientation, a new method of ordering biological laws. Our answer is that it would indeed be useful if by means of it we could formulate more general laws than could be formulated before. In physics, for example, the concept of energy played such a role. Nineteenth-century physicists theorized that perhaps certain phenomena, such as kinetic and potential energy in mechanics, heat (this was before the discovery that heat is simply the kinetic energy of molecules), the energy of magnetic fields, and so on, might be manifestations of one basic kind of energy. This led to experiments showing that mechanical energy can be transformed into heat and heat into mechanical energy but that the amount of energy remains constant. Thus, energy was a fruitful concept because it led to more general laws, such as the law of the conservation of energy. But Driesch's entelechy was not a fruitful concept in this sense. It did not lead to the discovery of more general biological laws.

In addition to providing *explanations* for observed facts, the laws of science also provide a means for *predicting* new facts not yet observed. The logical schema involved here is exactly the same as the schema underlying explanation. . . . Expressed symbolically:

$$1 \ (x) (Px \supset Qx)$$

$$2 \ Pa$$

$$3 \ Qa$$

First we have a universal law: for any object x , if it has the property P , then it also has the property Q . Second, we have a statement saying that object a has the property P . Third, we deduce by elementary logic that object a has the property Q . This schema underlies both explanation and prediction; only the knowledge situation is different. In explanation, the fact Qa is already known. We explain Qa by showing how it can be deduced from statements 1 and 2. In prediction, Qa is a fact *not yet known*. We have a law, and we have the fact Pa . We conclude that Qa must also

be a fact, even though it has not yet been observed. For example, I know the law of thermal expansion. I also know that I have heated a certain rod. By applying logic in the way shown in the schema, I infer that if I now measure the rod, I will find that it is longer than it was before.

In most cases, the unknown fact is actually a future event (for example, an astronomer predicts the time of the next eclipse of the sun); that is why I use the term "prediction" for this second use of laws. It need not, however, be prediction in the literal sense. In many cases the unknown fact is simultaneous with the known fact, as is the case in the example of the heated rod. The expansion of the rod occurs simultaneously with the heating. It is only our observation of the expansion that takes place after our observation of the heating.

In other cases, the unknown fact may even be in the past. On the basis of psychological laws, together with certain facts derived from historical documents, a historian infers certain unknown facts of history. An astronomer may infer that an eclipse of the moon must have taken place at a certain date in the past. A geologist may infer from striations on boulders that at one time in the past a region must have been covered by a glacier. I use the term "prediction" for all these examples because in every case we have the same logical schema and the same knowledge situation—a known fact and a known law from which an unknown fact is derived.

In many cases, the law involved may be statistical rather than universal. The prediction will then be only probable. A meteorologist, for instance, deals with a mixture of exact physical laws and various statistical laws. He cannot say that it will rain tomorrow; he can only say that rain is very likely.

This uncertainty is also characteristic of prediction about human behavior. On the basis of knowing certain psychological laws of a statistical nature and certain facts about a person, we can predict with varying degrees of probability how he will behave. Perhaps we ask a psychologist to tell us what effect a certain event will have on our child. He replies: "As I see the situation, your child will probably react in this way. Of course, the laws of psychology are not very exact. It is a young science, and as yet we know very little about its laws. But on the basis of what is known, I think it advisable that you plan to . . .". And so he gives us advice based on the best prediction he can make, with his probabilistic laws, about the future behavior of our child.

When the law is universal, then elementary deductive logic is involved in inferring unknown facts. If the law is statistical, we must use a different logic—the logic of probability. To give a simple example: a law states that 90 per cent of the residents of a certain region have black hair. I know that an individual is a resident of that region, but I do not know the color of his hair. I can infer, however, on the basis of the statistical law, that the probability his hair is black is $\frac{9}{10}$.

Prediction is, of course, as essential to everyday life as it is to science. Even the most trivial acts we perform during the day are based on predictions. You turn a doorknob. You do so because past observations of facts, together with universal laws, lead you to believe that turning the knob will open the door. You may not be conscious of the logical schema involved—no doubt you are thinking about other things—but all such deliberate actions presuppose the schema. There is a knowledge of specific facts, a knowledge of certain observed regularities that can be expressed as universal or statistical laws and provide a basis for the prediction of unknown facts. Prediction is involved in every act of human behavior that involves deliberate choice. Without it, both science and everyday life would be impossible.

Two Basic Types of Scientific Explanation

1 | Deductive-Nomological Explanation

In his book, *How We Think*,¹ John Dewey describes an observation he made one day when, washing dishes, he took some glass tumblers out of the hot soap suds and put them upside down on a plate: he noticed that soap bubbles emerged from under the tumblers' rims, grew for a while, came to a standstill, and finally receded inside the tumblers. Why did this happen? The explanation Dewey outlines comes to this: In transferring a tumbler to the plate, cool air is caught in it; this air is gradually warmed by the glass, which initially has the temperature of the hot suds. The warming of the air is accompanied by an increase in its pressure, which in turn produces an expansion of the soap film between the plate and the rim. Gradually, the glass cools off, and so does the air inside, with the result that the soap bubbles recede.

This explanatory account may be regarded as an argument to the effect that the event to be explained (let me call it the explanandum-event) was to be expected by reason of certain explanatory facts. These may be divided into two groups: (i) particular facts and (ii) uniformities expressed by general laws. The first group includes facts such as these: the tumblers had been immersed for some time, in soap suds of a temperature considerably higher than that of the surrounding air; they were put, upside down, on a plate on which a puddle of soapy water had formed, providing a connecting soap film, etc. The second group of items presupposed in the argument includes the gas laws and various other laws that have not been explicitly suggested concerning the exchange of heat between bodies of different temperature, the elastic behavior of soap bubbles, etc. If we imag-

FROM "Explanation in Science and History," in *Frontiers of Science and Philosophy*, ed. R. G. Colodny (London and Pittsburgh: Allen and Unwin and University of Pittsburgh Press, 1962), 9-19, 32.

ine these various presuppositions explicitly spelled out, the idea suggests itself of construing the explanation as a deductive argument of this form:

$$\begin{array}{l} \text{(D)} \quad C_1, C_2, \dots, C_k \\ \quad \quad L_1, L_2, \dots, L_r \\ \hline \quad \quad E \end{array}$$

Here, C_1, C_2, \dots, C_k are statements describing the particular facts invoked; L_1, L_2, \dots, L_r are general laws: jointly, these statements will be said to form the explanans. The conclusion E is a statement describing the explanandum-event; let me call it the explanandum-statement, and let me use the word "explanandum" to refer to either E or to the event described by it.

The kind of explanation thus characterized I will call *deductive-nomological explanation*; for it amounts to a deductive subsumption of the explanandum under principles which have the character of general laws: it answers the question "Why did the explanandum event occur?" by showing that the event resulted from the particular circumstances specified in C_1, C_2, \dots, C_k in accordance with the laws L_1, L_2, \dots, L_r . This conception of explanation, as exhibited in schema (D), has therefore been referred to as the covering law model, or as the deductive model, of explanation.^{2*}

A good many scientific explanations can be regarded as deductive-nomological in character. Consider, for example, the explanation of mirror-images, of rainbows, or of the appearance that a spoon handle is bent at the point where it emerges from a glass of water: in all these cases, the explanandum is deductively subsumed under the laws of reflection and refraction. Similarly, certain aspects of free fall and of planetary motion can be accounted for by deductive subsumption under Galileo's or Kepler's laws.

In the illustrations given so far the explanatory laws had, by and large, the character of empirical generalizations connecting different observable aspects of the phenomena under scrutiny: angle of incidence with angle of reflection or refraction, distance covered with falling time, etc. But science raises the question "why?" also with respect to the uniformities expressed by such laws, and often answers it in basically the same manner, namely, by subsuming the uniformities under more inclusive laws, and eventually under comprehensive theories. For example, the question, "Why do Galileo's and Kepler's laws hold?" is answered by showing that these laws are but special consequences of the Newtonian laws of motion

* The phrase *covering law model* is appropriate because general laws must "cover" or subsume the explanandum. The adjective *nomological* in the phrase *deductive-nomological* is derived from the Greek word *nomos*, meaning "law."

and of gravitation; and these, in turn, may be explained by subsumption under the more comprehensive general theory of relativity. Such subsumption under broader laws or theories usually increases both the breadth and the depth of our scientific understanding. There is an increase in breadth, or scope, because the new explanatory principles cover a broader range of phenomena; for example, Newton's principles govern free fall on the earth and on other celestial bodies, as well as the motions of planets, comets, and artificial satellites, the movements of pendulums, tidal changes, and various other phenomena. And the increase thus effected in the depth of our understanding is strikingly reflected in the fact that, in the light of more advanced explanatory principles, the original empirical laws are usually seen to hold only approximately, or within certain limits. For example, Newton's theory implies that the factor g in Galileo's law, $s = \frac{1}{2} gt^2$, is not strictly a constant for free fall near the surface of the earth; and that, since every planet undergoes gravitational attraction not only from the sun, but also from the other planets, the planetary orbits are not strictly ellipses, as stated in Kepler's laws.

One further point deserves brief mention here. An explanation of a particular event is often conceived as specifying its *cause*, or causes. Thus, the account outlined in our first illustration might be held to explain the growth and the recession of the soap bubbles by showing that the phenomenon was *caused* by a rise and a subsequent drop of the temperature of the air trapped in the tumblers. Clearly, however, these temperature changes provide the requisite explanation only in conjunction with certain other conditions, such as the presence of a soap film, practically constant pressure of the air surrounding the glasses, etc. Accordingly, in the context of explanation, a cause must be allowed to consist in a more or less complex set of particular circumstances; these might be described by a set of sentences: C_1, C_2, \dots, C_k . And, as suggested by the principle "Same cause, same effect," the assertion that those circumstances jointly caused a given event—described, let us say, by a sentence E —implies that whenever and wherever circumstances of the kind in question occur, an event of the kind to be explained comes about. Hence, the given causal explanation implicitly claims that there are general laws—such as L_1, L_2, \dots, L_i in schema (D)—by virtue of which the occurrence of the causal antecedents mentioned in C_1, C_2, \dots, C_k is a sufficient condition for the occurrence of the event to be explained. Thus, the relation between causal factors and effect is reflected in schema (D): causal explanation is deductive-nomological in character. (However, the customary formulations of causal and other explanations often do not explicitly specify all the relevant laws and particular facts: to this point, we will return later.)

The converse does not hold: there are deductive-nomological explanations which would not normally be counted as causal. For one thing, the subsumption of laws, such as Galileo's or Kepler's laws, under more comprehensive principles is clearly not causal in character: we speak of

causes only in reference to *particular* facts or events, and not in reference to *universal* facts as expressed by general laws. But not even all deductive-nomological explanations of particular facts or events will qualify as causal; for in a causal explanation some of the explanatory circumstances will temporally precede the effect to be explained: and there are explanations of type (D) which lack this characteristic. For example, the pressure which a gas of specified mass possesses at a given time might be explained by reference to its temperature and its volume at the same time, in conjunction with the gas law which connects simultaneous values of the three parameters.³

In conclusion, let me stress once more the important role of laws in deductive-nomological explanation: the laws connect the explanandum event with the particular conditions cited in the explanans, and this is what confers upon the latter the status of explanatory (and, in some cases, causal) factors in regard to the phenomenon to be explained.

2 | Probabilistic Explanation

In deductive-nomological explanation as schematized in (D), the laws and theoretical principles involved are of *strictly universal form*: they assert that in *all* cases in which certain specified conditions are realized an occurrence of such and such a kind will result; the law that any metal, when heated under constant pressure, will increase in volume, is a typical example; Galileo's, Kepler's, Newton's, Boyle's, and Snell's laws, and many others, are of the same character.

Now let me turn next to a second basic type of scientific explanation. This kind of explanation, too, is nomological, i.e., it accounts for a given phenomenon by reference to general laws or theoretical principles; but some or all of these are of *probabilistic-statistical form*, i.e., they are, generally speaking, assertions to the effect that if certain specified conditions are realized, then an occurrence of such and such a kind will come about with such and such a statistical probability.

For example, the subsiding of a violent attack of hay fever in a given case might well be attributed to, and thus explained by reference to, the administration of 8 milligrams of chlor-trimeton. But if we wish to connect this antecedent event with the explanandum, and thus to establish its explanatory significance for the latter, we cannot invoke a universal law to the effect that the administration of 8 milligrams of that antihistamine will invariably terminate a hay fever attack: this simply is not so. What can be asserted is only a generalization to the effect that administration of the drug will be followed by relief with high statistical probability, i.e., roughly speaking, with a high relative frequency in the long run. The resulting explanans will thus be of the following type:

John Doe had a hay fever attack and took 8 milligrams of chlor-trimeton.

The probability for subsidence of a hay fever attack upon administration of 8 milligrams of chlor-trimeton is high.

Clearly, this explanans does not deductively imply the explanandum, "John Doe's hay fever attack subsided"; the truth of the explanans makes the truth of the explanandum not certain (as it does in a deductive-nomological explanation) but only more or less likely or, perhaps "practically" certain.

Reduced to its simplest essentials, a probabilistic explanation thus takes the following form:

$$(P) \left. \begin{array}{l} F_i \\ \underline{p(O,F) \text{ is very high}} \\ O_i \end{array} \right\} \text{ makes very likely}$$

The explanandum, expressed by the statement " O_i ," consists in the fact that in the particular instance under consideration, here called i (e.g., John Doe's allergic attack), an outcome of kind O (subsidence) occurred. This is explained by means of two explanans-statements. The first of these, " F_i ," corresponds to C_1, C_2, \dots, C_n in (D); it states that in case i , the factors F (which may be more or less complex) were realized. The second expresses a law of probabilistic form, to the effect that the statistical probability for outcome O to occur in cases where F is realized is very high (close to 1). The double line separating explanandum from explanans is to indicate that, in contrast to the case of deductive-nomological explanation, the explanans does not logically imply the explanandum, but only confers a high likelihood upon it. The concept of likelihood here referred to must be clearly distinguished from that of statistical probability, symbolized by " p " in our schema. A statistical probability is, roughly speaking, the long-run relative frequency with which an occurrence of a given kind (say, F) is accompanied by an "outcome" of a specified kind (say, O). Our likelihood, on the other hand, is a relation (capable of gradations) not between kinds of occurrences, but between statements. The likelihood referred to in (P) may be characterized as the strength of the inductive support, or the degree of rational credibility, which the explanans confers upon the explanandum; or, in Carnap's terminology, as the *logical*, or *inductive*, (in contrast to statistical) *probability* which the explanandum possesses relative to the explanans.

Thus, probabilistic explanation, just like explanation in the manner of schema (D), is nomological in that it presupposes general laws; but because these laws are of statistical rather than of strictly universal form, the resulting explanatory arguments are inductive rather than deductive

in character. An inductive argument of this kind *explains* a given phenomenon by showing that, in view of certain particular events and certain statistical laws, its occurrence was to be expected with high logical, or inductive, probability.

By reason of its inductive character, probabilistic explanation differs from its deductive-nomological counterpart in several other important respects; for example, its explanans may confer upon the explanandum a more or less high degree of inductive support; in this sense, probabilistic explanation admits of degrees, whereas deductive-nomological explanation appears as an either-or affair: a given set of universal laws and particular statements either does or does not imply a given explanandum statement. A fuller examination of these differences, however, would lead us far afield and is not required for the purposes of this paper.*

One final point: the distinction here suggested between deductive-nomological and probabilistic explanation might be questioned on the ground that, after all, the universal laws invoked in a deductive explanation can have been established only on the basis of a finite body of evidence, which surely affords no exhaustive verification, but only more or less strong probability for it; and that, therefore, all scientific laws have to be regarded as probabilistic. This argument, however, confounds a logical issue with an epistemological one: it fails to distinguish properly between the *claim* made by a given law-statement and the *degree of confirmation*, or *probability*, which it possesses on the available evidence. It is quite true that statements expressing laws of either kind can be only incompletely confirmed by any given finite set—however large—of data about particular facts; but law-statements of the two different types make claims of different kind, which are reflected in their logical forms: roughly, a universal law-statement of the simplest kind asserts that *all* elements of an indefinitely large reference class (e.g., copper objects) have a certain characteristic (e.g., that of being good conductors of electricity); while statistical law-statements assert that in the long run, a specified proportion of the members of the reference class have some specified property. And our distinction of two types of law and, concomitantly, of two types of scientific explanation, is based on this difference in claim as reflected in the difference of form.

The great scientific importance of probabilistic explanation is eloquently attested to by the extensive and highly successful explanatory use that has been made of fundamental laws of statistical form in genetics, statistical mechanics, and quantum theory.

3 | Elliptic and Partial Explanations: Explanation Sketches

As I mentioned earlier, the conception of deductive-nomological explanation reflected in our schema (D) is often referred to as the covering law model, or the deductive model, of explanation; similarly, the conception underlying schema (P) might be called the probabilistic or the inductive-statistical, model of explanation. The term "model" can serve as a useful reminder that the two types of explanation as characterized above constitute ideal types or theoretical idealizations and are not intended to reflect the manner in which working scientists actually formulate their explanatory accounts. Rather, they are meant to provide explications, or rational reconstructions, or theoretical models, of certain modes of scientific explanation.

In this respect our models might be compared to the concept of mathematical proof (within a given theory) as construed in meta-mathematics. This concept, too, may be regarded as a theoretical model: it is not intended to provide a descriptive account of how proofs are formulated in the writings of mathematicians: most of these actual formulations fall short of rigorous and, as it were, ideal, meta-mathematical standards. But the theoretical model has certain other functions: it exhibits the rationale of mathematical proofs by revealing the logical connections underlying the successive steps; it provides standards for a critical appraisal of any proposed proof constructed within the mathematical system to which the model refers; and it affords a basis for a precise and far-reaching theory of proof, provability, decidability, and related concepts. I think the two models of explanation can fulfill the same functions, if only on a much more modest scale. For example, the arguments presented in constructing the models give an indication of the sense in which the models exhibit the rationale and the logical structure of the explanations they are intended to represent.

I now want to add a few words concerning the second of the functions just mentioned, but I will have to forgo a discussion of the third.

When a mathematician proves a theorem, he will often omit mention of certain propositions which he presupposes in his argument and which he is in fact entitled to presuppose because, for example, they follow readily from the postulates of his system or from previously established theorems or perhaps from the hypothesis of his theorem, if the latter is in hypothetical form; he then simply assumes that his readers or listeners will be able to supply the missing items if they so desire. If judged by ideal standards, the given formulation of the proof is elliptic or incomplete; but the departure from the ideal is harmless: the gaps can readily be filled in. Similarly, explanations put forward in everyday discourse and also in scientific contexts are often *elliptically formulated*. When we explain, for

example, that a lump of butter melted because it was put into a hot frying pan, or that a small rainbow appeared in the spray of the lawn sprinkler because the sunlight was reflected and refracted by the water droplets, we may be said to offer elliptic formulations of deductive-nomological explanations; an account of this kind omits mention of certain laws or particular facts which it tacitly takes for granted, and whose explicit citation would yield a complete deductive-nomological argument.

In addition to elliptic formulation, there is another, quite important, respect in which many explanatory arguments deviate from the theoretical model. It often happens that the statement actually included in the explanans, together with those which may reasonably be assumed to have been taken for granted in the context at hand, explain the given explanandum only *partially*, in a sense which I will try to indicate by an example. In his *Psychopathology of Everyday Life*, Freud offers the following explanation of a slip of the pen that occurred to him: "On a sheet of paper containing principally short daily notes of business interest, I found, to my surprise, the incorrect date, 'Thursday, October 20th,' bracketed under the correct date of the month of September. It was not difficult to explain this anticipation as the expression of a wish. A few days before I had returned fresh from my vacation and felt ready for any amount of professional work, but as yet there were few patients. On my arrival I had found a letter from a patient announcing her arrival on the 20th of October. As I wrote the same date in September I may certainly have thought 'X ought to be here already; what a pity about that whole month!', and with this thought I pushed the current date a month ahead."

Clearly, the formulation of the intended explanation is *at least incomplete* in the sense considered a moment ago. In particular, it fails to mention any laws or theoretical principles in virtue of which the subconscious wish, and the other antecedent circumstances referred to, could be held to explain Freud's slip of the pen. However, the general theoretical considerations Freud presents here and elsewhere in his writings suggests strongly that his explanatory account relies on a hypothesis to the effect that when a person has a strong, though perhaps unconscious, desire, then if he commits a slip of pen, tongue, memory, or the like, the slip will take a form in which it expresses, and perhaps symbolically fulfills, the given desire.

Even this rather vague hypothesis is probably more definite than what Freud would have been willing to assert. But for the sake of the argument let us accept it and include it in the explanans, together with the particular statements that Freud did have the subconscious wish he mentions, and that he was going to commit a slip of the pen. Even then, the resulting explanans permits us to deduce only that the slip made by Freud would, *in some way or other*, express and perhaps symbolically fulfill Freud's subconscious wish. But clearly, such expression and fulfillment might have

been achieved by many other kinds of slip of the pen than the one actually committed.

In other words, the explanans does not imply, and thus fully explain, that the particular slip, say *s*, which Freud committed on this occasion, would fall within the narrow class, say *W*, of acts which consist in writing the words "Thursday, October 20th"; rather, the explanans implies only that *s* would fall into a wider class, say *F*, which includes *W* as a proper subclass, and which consists of all acts which would express and symbolically fulfill Freud's subconscious wish in some way or other.

The argument under consideration might be called a *partial explanation*: it provides complete, or conclusive, grounds for expecting *s* to be a member of *F*, and since *W* is a subclass of *F*, it thus shows that the explanandum, i.e., *s* falling within *W*, accords with, or bears out, what is to be expected in consideration of the explanans. By contrast, a deductive-nomological explanation of the form (D) might then be called *complete* since the explanans here does imply the explanandum.

Clearly, the question whether a given explanatory argument is complete or partial can be significantly raised only if the explanandum sentence is fully specified; only then can we ask whether the explanandum does or does not follow from the explanans. Completeness of explanation, in this sense, is relative to our explanandum sentence. Now, it might seem much more important and interesting to consider instead the notion of a complete explanation of some concrete event, such as the destruction of Pompeii, or the death of Adolf Hitler, or the launching of the first artificial satellite: we might want to regard a particular event as completely explained only if an explanatory account of deductive or of inductive form had been provided for all of its aspects. This notion, however, is self-defeating; for any particular event may be regarded as having infinitely many different aspects or characteristics, which cannot all be accounted for by a finite set, however large, of explanatory statements.

In some cases, what is intended as an explanatory account will depart even further from the standards reflected in the model schemata (D) and (P) above. An explanatory account, for example, which is not explicit and specific enough to be reasonably qualified as an elliptically formulated explanation or as a partial one, can often be viewed as an *explanation sketch*: it may suggest, perhaps quite vividly and persuasively, the general outlines of what, it is hoped, can eventually be supplemented so as to yield a more closely reasoned argument based on explanatory hypotheses which are indicated more fully, and which more readily permit of critical appraisal by reference to empirical evidence.

The decision whether a proposed explanatory account is to be qualified as an elliptically formulated deductive or probabilistic explanation, as a partial explanation, as an explanation sketch, or perhaps as none of these is a matter of judicious interpretation; it calls for an appraisal of the

intent of the given argument and of the background assumptions that may be assumed to have been tacitly taken for granted, or at least to be available, in the given context. Unequivocal decision rules cannot be set down for this purpose any more than for determining whether a given informally stated inference which is not deductively valid by reasonably strict standards is to count nevertheless as valid but enthymematically formulated, or as fallacious, or as an instance of sound inductive reasoning, or perhaps, for lack of clarity, as none of these. . . .

■ | Notes

1. See Dewey, John. *How We Think*. Boston, New York, Chicago, 1910; Chapter VI.
2. For a fuller presentation of the model and for further references, see, for example, Hempel, C. G. and P. Oppenheim, "Studies in the Logic of Explanation," *Philosophy of Science* 15: 135-175 (1948). (Secs. 1-7 of this article, which contain all the fundamentals of the presentation, are reprinted in Feigl, H. and M. Scriven (eds.), *Readings in the Philosophy of Science*. New York, 1953.)—The suggestive term "covering law model" is W. Dray's; cf. his *Laws and Explanation in History*. Oxford, 1957; Chapter I. Dray characterizes this type of explanation as "subsuming what is to be explained under a general law" (*loc. cit.*, p. 1), and then rightly urges, in the name of methodological realism, that "the requirement of a single law be dropped" (*loc. cit.*, p. 24; italics, the author's); it should be noted, however, that, like the schema (D) above, several earlier publications on the subject (among them the article mentioned at the beginning of this note) make explicit provision for the inclusion of more laws than one in the explanans.
3. The relevance of the covering-law model to causal explanation is examined more fully in sec. 4 of Hempel, C. G., "Deductive-Nomological vs. Statistical Explanation." In Feigl, H., et al. (eds.), *Minnesota Studies in the Philosophy of Science*. vol. III. Minneapolis, 1962.
4. The concept of probabilistic explanation, and some of the peculiar logical and methodological problems engendered by it, are examined in some detail in Part II of the essay cited in note 3.
5. Freud, S. *Psychopathology of Everyday Life*. Translated by A. A. Brill. New York (Mentor Books) 1951; p. 64.

The Thesis of Structural Identity

Since in a fully stated D-N explanation of a particular event the explanans logically implies the explanandum, we may say that the explanatory argument might have been used for a deductive prediction of the explanandum-event if the laws and the particular facts adduced in its explanans had been known and taken into account at a suitable earlier time. In this sense, a D-N explanation is a potential D-N prediction.

This point was made already in an earlier article by Oppenheim and myself,¹ where we added that scientific explanation (of the deductive-nomological kind) differs from scientific prediction not in logical structure, but in certain pragmatic respects. In one case, the event described in the conclusion is known to have occurred, and suitable statements of general law and particular fact are sought to account for it; in the other, the latter statements are given and the statement about the event in question is derived from them before the time of its presumptive occurrence. This conception, which has sometimes been referred to as the *thesis of the structural identity* (or of the symmetry) of explanation and prediction, has recently been questioned by several writers. A consideration of some of their arguments may help to shed further light on the issues involved.

To begin with, some writers² have noted that what is usually called a prediction is not an argument but a sentence. More precisely, as Scheffler has pointed out, it is a sentence-token, i.e., a concrete utterance or inscription of a sentence purporting to describe some event that is to occur after the production of the token.³ This is certainly so. But in empirical science predictive sentences are normally established on the basis of available information by means of arguments that may be deductive or inductive in character; and the thesis under discussion should be understood, of course, to refer to explanatory and predictive arguments.

Thus construed, the *thesis of structural identity* amounts to the con-

junction of two sub-theses, namely (i) that every adequate explanation is potentially a prediction in the sense indicated above; (ii) that conversely every adequate prediction is potentially an explanation. I will now examine a number of objections that have been raised against the thesis, dealing first with those which, in effect, concern the first sub-thesis, and then with those concerning the second sub-thesis. I will argue that the first sub-thesis is sound, whereas the second one is indeed open to question. Though the following considerations are concerned principally with D-N explanation, some of them are applicable to other types of explanation as well. . . .

The first sub-thesis, as has already been noted, is an almost trivial truth in the case of D-N explanation, since here the explanans logically implies the explanandum. But it is supported also by a more general principle, which applies to other types of explanation as well, and which expresses, I would submit, a general condition of adequacy for any rationally acceptable explanation of a particular event. That condition is the following: Any rationally acceptable answer to the question 'Why did event X occur?' must offer information which shows that X was to be expected—if not definitely, as in the case of D-N explanation, then at least with reasonable probability. Thus, the explanatory information must provide good grounds for believing that X did in fact occur; otherwise, that information would give us no adequate reason for saying: "That explains it—that does show why X occurred." And an explanatory account that satisfies this condition constitutes, of course, a potential prediction in the sense that it could have served to predict the occurrence of X (deductively or with more or less high probability) if the information contained in the explanans had been available at a suitable earlier time.

The condition of adequacy just stated can be extended, in an obvious manner, to explanations concerned, not with individual events, but with empirical uniformities expressed by putative laws. But such explanations cannot well be spoken of as potential predictions since law-statements purport to express timeless uniformities and thus make no reference to any particular time, whether past, present, or future.⁴

It will hardly be necessary to emphasize that it is not, of course, the purpose of an explanation to provide grounds in support of the explanandum-statement; for . . . a request for an explanation normally presupposes that the explanandum-statement is true. The point of the preceding remarks is rather that an adequate explanation cannot help providing information which, if properly established, also provides grounds in support of the explanandum-statement. . . . We may say that an adequate answer to an explanation-seeking why-question is always also a potential answer to the corresponding epistemic why-question.

The converse, however, does not hold; the condition of adequacy is necessary but not sufficient for an acceptable explanation. For example, certain empirical findings may give excellent grounds for the belief that the orientation of the earth's magnetic field shows diurnal and secular

variations, without in the least explaining why. Similarly, a set of experimental data may strongly *support* the assumption that the electric resistance of metals increases with their temperature or that a certain chemical inhibits the growth of cancer cells, without providing any *explanation* for these presumptive empirical regularities. The predictive inferences here involved are inductive rather than deductive; but what bars them from the status of potential explanations is not their inductive character . . . , but the fact that they invoke no laws or theoretical principles, no explanatory statements that make a general claim. Reliance on general principles, while perhaps not indispensable for prediction, is required in any explanation: such principles alone can give to whatever particular circumstances may be adduced the status of explanatory factors for the event to be explained.

Some of the objections recently raised against the thesis of the structural identity of explanation and prediction concern in effect the first of its two sub-theses, which has now been presented in some detail: the claim that any adequate explanatory argument is also potentially predictive. I will consider three objections to the effect that there are certain perfectly satisfactory explanations that do not constitute potential predictions.

Scriven has argued that the occurrence of an event *X* is sometimes quite adequately explained by means of a "proposition of the form 'The only cause of *X* is *A*' . . . for example, 'The only cause of paresis is syphilis';" this proposition enables us to explain why a certain patient has paresis by pointing out that he previously suffered from syphilis. And this explanation holds good, according to Scriven, even though only quite a small percentage of syphilitic patients develop paresis, so that "we must, on the evidence [that a given person has syphilis], still predict that [paresis] will not occur."⁵ But if it does occur, then the principle that the only cause of paresis is syphilis can "provide and guarantee our explanation" in terms of antecedent syphilitic infection.⁶ Thus we have here a presumptive explanation which indeed is not adequate as a potential prediction. But precisely because paresis is such a rare sequel of syphilis, prior syphilitic infection surely cannot by itself provide an adequate explanation for it. A condition that is *nominally* necessary for the occurrence of an event does not, in general, explain it; or else we would be able to explain a man's winning the first prize in the Irish sweepstakes by pointing out that he had previously bought a ticket, and that only a person who owns a ticket can win the first prize.

A second argument which, like Scriven's, has considerable initial plausibility has been advanced by Toulmin⁷ by reference to "Darwin's theory, explaining the origin of species by variation and natural selection. No scientist has ever used this theory to foretell the coming-into-existence of creatures of a novel species, still less verified his forecast. Yet many competent scientists have accepted Darwin's theory as having great explanatory power." In examining this argument, let me distinguish what

might be called the story of evolution from the *theory* of the underlying mechanisms of mutation and natural selection. The story of evolution, as a hypothesis about the gradual development of various types of organisms, and about the subsequent extinction of many of these, has the character of a hypothetical historical narrative *describing* the putative stages of the evolutionary process: it is the associated theory which provides what *explanatory insight* we have into this process. The story of evolution might tell us, for example, that at a certain stage in the process dinosaurs made their appearance and that, so much later, they died out. Such a narrative account does not, of course, explain why the various kinds of dinosaurs with their distinctive characteristics came into existence, nor does it explain why they became extinct. Indeed even the associated theory of mutation and natural selection does not answer the first of these questions, though it might be held to shed some light on the latter. Yet, even to account for the extinction of the dinosaurs, we need a vast array of additional hypotheses about their physical and biological environment and about the species with which they had to compete for survival. But if we have hypotheses of this kind that are specific enough to provide, in combination with the theory of natural selection, at least a probabilistic explanation for the extinction of the dinosaurs, then clearly the explanans adduced is also qualified as a basis for a potential probabilistic prediction. The undeniably great persuasiveness of Toulmin's argument would seem to derive from two sources, a widespread tendency to regard the basically descriptive story of evolution as explaining the various states of the process, and a similarly widespread tendency to overestimate the extent to which even the theory of mutation and natural selection can account for the details of the evolutionary sequence.

I now turn to a third objection to the claim that an adequate explanation is also a potential prediction. It is based on the observation that sometimes the only ground we have for asserting some essential statement in the explanans lies in the knowledge that the explanandum event *did* in fact occur. In such cases, the explanatory argument clearly could not have been used to predict that event. Consider one of Scriven's examples.⁶ Suppose that a man has killed his wife whom he knew to have been unfaithful to him, and that his action is explained as the result of intense jealousy. The fact that the man was jealous might well have been ascertainable before the deed, but to explain the latter, we need to know that his jealousy was intense enough to drive him to murder; and this we can know only after the deed has actually been committed. Here then, the occurrence of the explanandum event provides the only grounds we have for asserting one important part of the explanans; the explanandum event could not therefore have been predicted by means of the explanatory argument. In another example,⁷ Scriven considers an explanation to the effect that the collapse of a bridge was caused by metal fatigue. This account, he argues, might be supported by pointing out that the failure could

have been caused only by an excessive load, by external damage, or by metal fatigue, and that the first two factors were not present in the case at hand, whereas there is evidence of metal fatigue. *Given the information that the bridge did in fact collapse*, this would establish not only that metal fatigue was at fault but that it was strong enough to cause the failure. While Scriven's notion of "the only possible cause" of a given event surely requires further elucidation, his example does afford another illustration of an explanatory account one of whose constituent hypotheses is supported only by the occurrence of the event to be explained—so that the latter could not have been predicted by means of the explanatory argument.

However, the point thus illustrated does not affect at all the conditional thesis that an adequate explanatory argument must be such that it could have served to predict the explanandum event *if* the information included in the explanans had been known and taken into account before the occurrence of that event. What Scriven's cases show is that sometimes we do not know independently of the occurrence of the explanandum event that all the conditions listed in the explanans are realized. However, this means only that in such cases our conditional thesis is counterfactual, i.e., that its if-clause is not satisfied, but not that the thesis itself is false. Moreover, Scriven's argument does not even show that in the kind of case he mentions it is logically or nomologically impossible (impossible by reason of the laws of logic or the laws of nature) for us to know the critical explanatory factor before, or independently of, the occurrence of the explanandum-event; the impossibility appears to be rather a practical and perhaps temporary one, reflecting present limitations of knowledge or technology.

But while it thus leaves our thesis unaffected, Scriven's observation is of methodological interest in its own right: it shows that sometimes an event is explained by means of hypotheses for some of which the fact of its occurrence affords the only available evidential support. This may happen, as we saw, when one of the explanatory hypotheses states that a certain relevant factor was strong enough to bring about the event in question; but the observation applies also to other cases. Thus the explanation . . . of the appearance and initial growth of the soap bubbles, includes in its explanans the assumption that a soap film had formed between the plate and the rims of the tumblers;^{*} and practically the only evidence available in support of this explanatory assumption is the fact that soap bubbles did emerge from under the tumblers. Or consider the explanation of the characteristic dark lines in the absorption spectrum of a particular star. The key assumption in the explanans is that the star's atmosphere

* See the first section of Hempel's "Two Basic Types of Scientific Explanation," (the preceding reading in this chapter) for more about why bubbles appear when a tumbler is taken from warm soapy water and inverted on a plate.

contains certain elements, such as hydrogen, helium, and calcium, whose atoms absorb radiation of the wave lengths corresponding to the dark lines; the explanation relies, of course, on many other assumptions, including the optical theory that forms the basis for spectroscopy, and the assumption that the apparatus used is a properly constructed spectroscope. But while these latter explanans statements are capable of independent test and corroboration, it may well be that the only evidence available in support of the key explanatory hypothesis is the occurrence of the very lines whose appearance in the spectrum the argument serves to explain. Strictly speaking, the explanandum event here provides support for the key explanatory hypothesis only by virtue of the background theory, which connects the presence of certain elements in the atmosphere of a star with the appearance of corresponding absorption lines in its spectrum. Thus, the information that the explanandum event has occurred does not by itself support the explanatory hypothesis in question, but it constitutes, as we might say, an essential part of the only evidence available in support of that hypothesis.

Explanations of the kind here considered may be schematically characterized as arguments of the form (D-N) in which the information or assumption that *E* is true provides an indispensable part of the only available evidential support for one of the explanans statements, say, *C*₁. Let us call such explanations *self-evidencing*. It might be held that the actual occurrence of the explanandum event always provides some slight additional support even for an explanans whose constituent sentences have been accepted on the basis of independent evidence, and that in this sense every D-N explanation with true explanandum is in some measure self-evidencing; but we will apply this appellation to an explanatory account only if, at the time of its presentation, the occurrence of the explanandum event provides the only evidence, or an indispensable part of the only evidence, available in support of some of the explanans-statements.

An explanatory argument of the form (D-N) which is self-evidencing is not for that reason circular or pointless. To be sure, if the same argument were adduced in support of the assertion that the explanandum-event did occur (or, that *E* is true), then it would be open to the charge of epistemic circularity. If the argument is to achieve its objective then all the grounds it adduces in support of *E*—i.e., *C*₁, *C*₂, . . . , *C*_{*n*}; *L*₁, *L*₂, . . . , *L*_{*r*}—would have to be established independently of *E*; and this condition is violated here since the only ground we have for believing or asserting *C*₁ includes the assumption that *E* is true. But when the same argument is used for explanatory purposes it does not claim to establish that *E* is true; that is *presupposed* by the question "Why did the event described by *E* occur?". Nor need a self-evidencing explanation involve an explanatory circle. The information that the explanandum event has occurred is not included in the explanans (so that the occurrence of the event is not "explained by

itself"); rather it serves, quite outside the explanatory context, as evidence supporting one of the explanans statements. Thus, an acceptable self-evidencing explanation benefits, as it were, by the wisdom of hindsight derived from the information that the explanandum event has occurred but it does not misuse that information so as to produce a circular explanation.

An explanation that is self-evidencing may for that reason rest on poorly supported explanans and may therefore have no strong claim to empirical soundness. But even this is not inevitable. In the case of the absorption spectrum of a star, for example, the previously accepted background information, including the relevant theories, may indicate that the dark lines observed occur *only* if the specified elements are present in the star's atmosphere; and then the explanandum, in conjunction with the background information, lends very strong support to the crucial explanatory hypothesis.

The notion of a self-evidencing explanation can, I think, shed some further light on the puzzle illustrated by the explanation of paresis in terms of antecedent syphilitic infection. Consider another illustration. Some cases of skin cancer are attributed to intensive ultraviolet irradiation. But this factor very often does not lead to cancer, so that the information that a person has been exposed to such radiation does not permit the prediction of cancer. Is that information alone nevertheless sufficient to explain the development of skin cancer when it does follow intensive irradiation? No doubt, an explanation will often be formulated so as to mention only the antecedent irradiation; but the underlying rationale surely must be more complex. Leaving aside the important quantitative aspects of the problem, the crucial point in that rationale can, I suggest, be schematically stated as follows: Some, though by no means all, individuals have the disposition to develop skin cancer upon exposure to strong ultraviolet irradiation; let us call these radiation-sensitive. Now, in the case of explanation, we know that the given individual was exposed to strong radiation (C_1) and did develop cancer of the skin in the affected area (E). But jointly, these two pieces of information lend support to the assumption that the individual is radiation-sensitive (C_2)—an hypothesis that is not supported in the case of prediction, where C_1 is available, but not E . And the two statements C_1 and C_2 (in combination with the general statement that sensitive individuals will develop skin cancer when exposed to intensive radiation) can provide an adequate explanans for E . If the explanation is thus constructed as invoking C_2 in addition to C_1 , it is seen to be self-evidencing, but also to possess an explanans which would provide an adequate basis for prediction if C_2 could be known in advance. That is impossible, of course, as long as the only available test for radiation-sensitivity consists in checking whether an individual does develop skin cancer upon intensive irradiation. But, clearly, it is conceivable that other, independent, tests

radiation-sensitivity might be found and then C_2 might well be established independently of, and even prior to, the occurrence of the event described by E .

In discussing the structural identity of explanation and prediction, I have so far considered only the first of the two sub-theses distinguished earlier, namely, the claim that every adequate explanation is also a potential prediction. I have argued that the objections raised against this claim fall short of their mark, and that the first sub-thesis is sound and can indeed serve as a necessary condition of adequacy for any explicitly stated, rationally acceptable explanation.

I turn now to the second sub-thesis, namely, that every adequate predictive argument also affords a potential explanation. This claim is open to question even in the case of certain predictive arguments that are of deductive-nomological character, as the following example illustrates. One of the early symptoms of measles is the appearance of small whitish spots, known as Koplik spots, on the mucous linings of the cheeks. The statement, L , that the appearance of Koplik spots is always followed by the later manifestations of the measles might therefore be taken to be a law, and it might then be used as a premise in D-N arguments with a second premise of the form 'Patient i has Koplik spots at time t ', and with a conclusion stating that i subsequently shows the later manifestations of the measles. An argument of this type is adequate for predictive purposes, but its explanatory adequacy might be questioned. We would not want to say, for example, that i had developed high fever and other symptoms of the measles because he had previously had Koplik spots. Yet this case—and others similar to it—does not constitute a decisive objection against the second sub-thesis. For the reluctance to regard the appearance of Koplik spots as explanatory may well reflect doubts as to whether, as a matter of universal law, those spots are always followed by the later manifestations of measles. Perhaps a local inoculation with a small amount of measles virus would produce the spots without leading to a full-blown case of the measles. If this were so, the appearance of the spots would still afford a usually reliable basis for predicting the occurrence of further symptoms, since exceptional conditions of the kind just mentioned would be extremely rare; but the generalization that Koplik spots are always followed by later symptoms of the measles would not express a law and thus could not properly support a corresponding D-N explanation.

The objection just considered concerns the explanatory potential of predictive arguments of the form (D-N). But the second sub-thesis, in its general form, which is not limited to D-N predictions, has further been challenged, particularly by Scheffler and by Scriven,¹⁰ on the ground that there are other kinds of predictive argument that are adequate for scientific prediction, yet not for explanation. Specifically, as Scheffler notes, a scientific prediction may be based on a finite set of data which includes no laws and which would have no explanatory force. For example, a finite

set of data obtained in an extensive test of the hypothesis that the electric resistance of metals increases with their temperature may afford good support for that hypothesis and may thus provide an acceptable basis for the prediction that in an as yet unexamined instance, a rise in temperature in a metal conductor will be accompanied by an increase in resistance. But if this event then actually occurs, the test data clearly do not provide an explanation for it. Similarly, a list of the results obtained in a long series of tossings of a given coin may provide a good basis for predicting the percentage of Heads and Tails to be expected in the next 1000 tossings of the same coin; but again, that list of data provides no explanation for the subsequent results. Cases like these raise the question of whether there are not sound modes of scientific prediction that proceed from particulars to particulars without benefit of general laws such as seem to be required for any adequate explanation. Now, the predictive arguments just considered are not deductive but probabilistic in character. . . . In regard to the second sub-thesis of the structural identity claim, let us note this much here: the predictions in our illustrations proceed from an observed sample of a population to another, as yet unobserved one; and on some current theories of probabilistic inference such arguments do not depend upon the assumption of general empirical laws. According to Carnap's theory of inductive logic,¹¹ for example, such inferences are possible on purely logical grounds; the information about the given sample confers a definite logical probability upon any proposed prediction concerning an as yet unobserved sample. On the other hand, certain statistical theories of probabilistic inference eschew the notion of purely logical probabilities and qualify predictions of the kind here considered as sound only on the further assumption that the selection of individual cases from the total population has the character of a random experiment with certain general statistical characteristics. But that assumption, when explicitly spelled out, has the form of a general law of statistic-probabilistic form; hence, the predictions are effected by means of covering laws after all. And though these laws do not have the strictly universal character of those invoked in D-N explanations and predictions, they can serve in an explanatory capacity as well. Thus construed, even the predictions here under discussion turn out to be (incompletely formulated) potential explanations. . . .

■ | Notes

1. Hempel and Oppenheim (1948), section 3.
2. See Scheffler (1957), section 1 and (1963), Part I, sections 3 and 4; Scriven (1962), p. 177.
3. Cf. Scheffler (1957), section 1. . . .
4. This point is made, for example, by Scriven (1962), pp. 179ff.

5. Scriven (1959a), p. 480, italics the author's.
6. Loc. cit. Barker has argued analogously that "it can be correct to speak of explanation in many cases where specific prediction is not possible. Thus, for instance, if the patient shows all the symptoms of pneumonia, sickens and dies, I can then explain his death—I know what killed him—but I could not have definitely predicted in advance that he was going to die; for usually pneumonia fails to be fatal." (1961, p. 271). This argument seems to me open to questions similar to those just raised in reference to Scriven's illustration. First of all, it is not clear just what would be claimed by the assertion that pneumonia killed the patient. Surely the mere information that the patient had pneumonia does not suffice to explain his death, precisely because in most cases pneumonia is not fatal. And if the explanans is taken to state that the patient was suffering from very severe pneumonia (and perhaps that he was elderly or weak) then it may well provide a basis at least for a probabilistic explanation of the patient's death—but in this case it obviously also permits prediction of his death with the same probability. For some further observations on Barker's argument, see the comments by Feyerabend and by Rudner, and Barker's rejoinders, in Feigl and Maxwell (1961), pp. 278–85. A detailed critical discussion that sheds further light on Scriven's paresis example will be found in Grünbaum (1963) and (1963a), chapter 9; see also Scriven's rejoinder (1963).
7. Toulmin (1961), pp. 24–25. Scriven (1959a) and Barker (1961) have offered arguments in the same vein. For a critical discussion of Scriven's version, see Grünbaum (1963) and (1963a), chapter 9.
8. Scriven (1959), pp. 468–69.
9. Scriven (1962), pp. 181–87.
10. See Scheffler (1957), p. 296 and (1963), p. 42; Scriven (1959a), p. 480.
11. Carnap (1950), section 110.

■ | Bibliography

- Barker, S. F., "The Role of Simplicity in Explanation," in Feigl and Maxwell (1961), 265–74.
- Baumrin, B. (ed.). *Philosophy of Science. The Delaware Seminar. Volume I, 1961–62* (New York: John Wiley & Sons, 1963).
- Carnap, R. *Logical Foundations of Probability* (Chicago: University of Chicago Press, 1950; 2nd revised ed. 1962).
- Feigl, H. and G. Maxwell (eds.). *Current Issues in the Philosophy of Science* (New York: Holt, Rinehart & Winston, 1961).
- Feigl, H. and G. Maxwell (eds.). *Minnesota Studies in the Philosophy of Science, Volume III* (Minneapolis: University of Minnesota Press, 1962).
- Gardiner, P. (ed.). *Theories of History* (New York: The Free Press, 1959).
- Grünbaum, A., "Temporally Asymmetric Principles, Parity between Explanations and Prediction, and Mechanism vs. Teleology," in Baumrin (1963), 57–96.

Grünbaum, A. *Philosophical Problems of Space and Time* (New York: Knopf, 1963a).

Hempel, C. G. and P. Oppenheim, "Studies in the Logic of Explanation," *Philosophy of Science* 15 (1948): 135-75.

Scheffler, I., "Explanation, Prediction, and Abstraction," *British Journal for the Philosophy of Science* 7 (1957), 293-309.

Scheffler, I. *The Anatomy of Inquiry: Philosophical Studies in the Theory of Science* (New York: Alfred A. Knopf, 1963).

Scriven, M., "Tautisms as the Grounds for Historical Explanations," in Gardiner (1959), 443-75.

Scriven, M., "Explanation and Prediction in Evolutionary Theory," *Science* 130 (1959a), 477-82.

Scriven, M., "Explanations, Predictions, and Laws," in Feigl and Maxwell (1962), 170-230.

Scriven, M., "The Temporal Asymmetry between Explanations and Predictions," in Baumrin (1963), 97-105.

Toulmin, S. *Foresight and Understanding* (London: Hutchinson, 1961; New York: Harper & Row, 1963).

Inductive-Statistical Explanation

1 | Inductive-Statistical Explanation

As an explanation of why patient John Jones recovered from a streptococcus infection, we might be told that Jones had been given penicillin. But if we try to amplify this explanatory claim by indicating a general connection between penicillin treatment and the subsiding of a streptococcus infection we cannot justifiably invoke a general law to the effect that in all cases of such infection, administration of penicillin will lead to recovery. What can be asserted, and what surely is taken for granted here, is only that penicillin will effect a cure in a high percentage of cases, or with a high statistical probability. This statement has the general character of a law of statistical form, and while the probability value is not specified, the statement indicates that it is high. But in contrast to the cases of deductive-nomological and deductive-statistical explanation, the explanans consisting of this statistical law together with the statement that the patient did receive penicillin obviously does not imply the explanandum statement, 'the patient recovered', with deductive certainty, but only, as we might say, with high likelihood, or near certainty. Briefly, then, the explanation amounts to this argument:

- 1a The particular case of illness of John Jones—let us call it j —was an instance of severe streptococcal infection (S_j) which was treated with large doses of penicillin (P_j); and the statistical probability $p(R, S \cdot P)$ of recovery in cases where S and P are present close to 1; hence, the case was practically certain to end in recovery (R_j).*

FROM *Aspects of Scientific Explanation* (New York: Free Press, 1965), 381–83, 394–403.

* Throughout this paper, Hempel uses a dot to stand for conjunction, a bar over a letter to stand for negation, and a comma within parentheses to represent conditional probabilities. Thus, for example, $p(R, S \cdot \bar{P})$ means the probability of R given S and not- P .

This argument might invite the following schematization:

1b $p(R, S \cdot P)$ is close to 1
 $S_j \cdot P_j$

(Therefore:) It is practically certain (very likely) that R_j

In the literature on inductive inference, arguments thus based on statistical hypotheses have often been construed as having this form or a similar one. On this construal, the conclusion characteristically contains a modal qualifier such as 'almost certainly,' 'with high probability,' 'very likely,' etc. But the conception of arguments having this character is untenable. For phrases of the form 'it is practically certain that p ' or 'It is very likely that p ', where the place of ' p ' is taken by some statement, are not complete self-contained sentences that can be qualified as either true or false. The statement that takes the place of ' p '—for example, ' R_j '—is either true or false, quite independently of whatever relevant evidence may be available, but it can be qualified as more or less likely, probable, certain, or the like only *relative to some body of evidence*. One and the same statement, such as ' R_j ', will be certain, very likely, not very likely, highly likely, and so forth, depending upon what evidence is considered. The phrase 'it is almost certain that R_j ' taken by itself is therefore neither true nor false; and it cannot be inferred from the premises specified in (1b) nor from any other statements.

The confusion underlying the schematization (1b) might be further illuminated by considering its analogue for the case of deductive arguments. The force of a deductive inference, such as that from 'all F are G ' and ' a is F ' to ' a is G ', is sometimes indicated by saying that if the premises are true, then the conclusion is necessarily true or is certain to be true—a phrasing that might suggest the schematization

All F are G
 a is F

(Therefore:) It is necessary (certain) that a is G

But clearly the given premises—which might be, for example, 'all men are mortal' and 'Socrates is a man'—do not establish the sentence ' a is G ' ('Socrates is mortal') as a necessary or certain truth. The certainty referred to in the informal paraphrase of the argument is relational: the statement ' a is G ' is certain, or necessary, *relative to the specified premises*; i.e., their truth will guarantee its truth—which means nothing more than that ' a is G ' is a logical consequence of those premises.

Analogously, to present our statistical explanation in the manner of schema (1b) is to misconstrue the function of the words 'almost certain' or 'very likely' as they occur in the formal wording of the explanation.

Those words clearly must be taken to indicate that on the evidence provided by the explanans, or relative to that evidence, the explanandum is practically certain or very likely, i.e., that

- 1c 'R_j' is practically certain (very likely) relative to the explanans containing the sentences ' $p(R, S \cdot P)$ is close to 1' and ' $S_j \cdot P_j$ '.¹

The explanatory argument misrepresented by (1b) might therefore suitably be schematized as follows:

$$\begin{array}{l} 1d \quad p(R, S \cdot P) \text{ is close to } 1 \\ \quad \underline{S_j \cdot P_j} \\ \quad R_j \end{array} \quad \text{[makes practically certain]}$$

In this schema, the double line separating the "premises" from the "conclusion" is to signify that the relation of the former to the latter is not that of deductive implication but that of inductive support, the strength of which is indicated in square brackets.² . . .

2 | The Problem of Explanatory Ambiguity

Consider once more the explanation (1d) of recovery in the particular case *j* of John Jones's illness. The statistical law there invoked claims recovery in response to penicillin only for a high percentage of streptococcal infections, but not for all of them; and in fact, certain streptococcus strains are resistant to penicillin. Let us say that an occurrence, e.g. a particular case of illness, has the property S^* (or belongs to the class S^*) if it is an instance of infection with a penicillin-resistant streptococcus strain. Then the probability of recovery among randomly chosen instances of S^* which are treated with penicillin will be quite small, i.e. $p(R, S^* \cdot P)$ will be close to 0 and the probability of nonrecovery, $p(\bar{R}, S^* \cdot P)$ will be close to 1. But suppose now that Jones's illness is in fact a streptococcal infection of the penicillin-resistant variety, and consider the following argument:

$$\begin{array}{l} 2a \quad p(\bar{R}, S^* \cdot P) \text{ is close to } 1 \\ \quad \underline{S^*_j \cdot P_j} \\ \quad \bar{R}_j \end{array} \quad \text{[makes practically certain]}$$

This "rival" argument has the same form as (1d), and on our assumptions, its premises are true, just like those of (1d). Yet its conclusion is the contradictory of the conclusion of (1d).

Or suppose that Jones is an octogenarian with a weak heart, and that

in this group, S^{**} , the probability of recovery from a streptococcus infection in response to penicillin treatment, $p(R, S^{**} \cdot P)$, is quite small. Then, there is the following rival argument to (1d), which presents Jones's nonrecovery as practically certain in the light of premises which are true:

$$2b \quad \frac{p(\bar{R}, S^{**} \cdot P) \text{ is close to } 1}{\frac{S^{**}j \cdot Pj}{Rj}} \quad \text{[makes practically certain]}$$

The peculiar logical phenomenon here illustrated will be called the *ambiguity of inductive-statistical explanation* or, briefly, of *statistical explanation*. This ambiguity derives from the fact that a given individual event (e.g., Jones's illness) will often be obtainable by random selection from any one of several "reference classes" (such as $S \cdot P$, $S^* \cdot P$, $S^{**} \cdot P$), with respect to which the kind of occurrence (e.g., R) instantiated by the given event has very different statistical probabilities. Hence, for a proposed probabilistic explanation with true explanans which confers near certainty upon a particular event, there will often exist a rival argument of the same probabilistic form and with equally true premises which confers near certainty upon the nonoccurrence of the same event. And any statistical explanation for the occurrence of an event must seem suspect if there is the possibility of a logically and empirically equally sound probabilistic account for its nonoccurrence. *This predicament has no analogue in the case of deductive explanation*; for if the premises of a proposed deductive explanation are true then so is its conclusion; and its contradictory, being false, cannot be a logical consequence of a rival set of premises that are equally true.

Here is another example of the ambiguity of I-S explanation: Upon expressing surprise at finding the weather in Stanford warm and sunny on a date as autumnal as November 27, I might be told, by way of explanation, that this was rather to be expected because the probability of warm and sunny weather (W) on a November day in Stanford (N) is, say, .95. Schematically, this account would take the following form, where 'n' stands for 'November 27':

$$2c \quad \frac{p(W, N) = .95}{\frac{Nn}{Wn}} \quad [.95]$$

But suppose it happens to be the case that the day before, November 26, was cold and rainy, and that the probability for the immediate successors (S) of cold and rainy days in Stanford to be warm and sunny is .2; then the account (2c) has a rival in the following argument which

by reference to equally true premises, presents it as fairly certain that November 27 is not warm and sunny:

$$2d \quad \frac{p(\overline{W}, S) = .8}{\frac{S_n}{W_n}} \quad [.8]$$

In this form, the problem of ambiguity concerns I-S arguments whose premises are in fact true, no matter whether we are aware of this or not. But, as will now be shown, the problem has a variant that concerns explanations whose explanans statements, no matter whether in fact true or not, are *asserted or accepted* by empirical science at the time when the explanation is proffered or contemplated. This variant will be called *the problem of the epistemic ambiguity of statistical explanation*, since it refers to what is presumed to be known in science rather than to what, perhaps unknown to anyone, is in fact the case.

Let K_t be the class of all statements asserted or accepted by empirical science at time t . This class then represents the total scientific information, or "scientific knowledge" at time t . The word 'knowledge' is here used in the sense in which we commonly speak of the scientific knowledge at a given time. It is not meant to convey the claim that the elements of K_t are true, and hence neither that they are definitely known to be true. No such claim can justifiably be made for any of the statements established by empirical science; and the basic standards of scientific inquiry demand that an empirical statement, however well supported, be accepted and thus admitted to membership in K_t only tentatively, i.e., with the understanding that the privilege may be withdrawn if unfavorable evidence should be discovered. The membership of K_t therefore changes in the course of time; for as a result of continuing research, new statements are admitted into that class; others may come to be discredited and dropped. Henceforth, the class of accepted statements will be referred to simply as K when specific reference to the time in question is not required. We will assume that K is logically consistent and that it is closed under logical implication, i.e., that it contains every statement that is logically implied by any of its subsets.

The *epistemic ambiguity of I-S explanation* can now be characterized as follows: The total set K of accepted scientific statements contains different subsets of statements which can be used as premises in arguments of the probabilistic form just considered, and which confer high probabilities on logically contradictory "conclusions." Our earlier examples (2a), (2b) and (2c), (2d) illustrate this point if we assume that the premises of those arguments all belong to K rather than that they are all true. If one of two such rival arguments with premises in K is proposed as an expla-

nation of an event considered, or acknowledged, in science to have occurred, then the conclusion of the argument, i.e., the explanandum statement, will accordingly belong to *K* as well. And since *K* is consistent, the conclusion of the rival argument will not belong to *K*. Nonetheless it is disquieting that we should be able to say: No matter whether we are informed that the event in question (e.g. warm and sunny weather on November 27 in Stanford) did occur or that it did not occur, we can produce an explanation of the reported outcome in either case; and an explanation, moreover, whose premises are scientifically established statements that confer a high logical probability upon the reported outcome.

This epistemic ambiguity, again, has no analogue for deductive explanation; for since *K* is logically consistent, it cannot contain premise-sets that imply logically contradictory conclusions.

Epistemic ambiguity also bedevils the predictive use of statistical arguments. Here, it has the alarming aspect of presenting us with two rival arguments whose premises are scientifically well established, but one of which characterizes a contemplated future occurrence as practically certain, whereas the other characterizes it as practically impossible. Which of such conflicting arguments, if any, are rationally to be relied on for explanation or for prediction?

3 | The Requirement of Maximal Specificity and the Epistemic Relativity of Inductive-Statistical Explanation

Our illustrations of explanatory ambiguity suggest that a decision on the acceptability of a proposed probabilistic explanation or prediction will have to be made in the light of all the relevant information at our disposal. This is indicated also by a general principle whose importance for inductive reasoning has been acknowledged, if not always very explicitly, by many writers, and which has recently been strongly emphasized by Carnap, who calls it *the requirement of total evidence*. Carnap formulates it as follows: "in the application of inductive logic to a given knowledge situation, the total evidence available must be taken as basis for determining the degree of confirmation." Using only a part of the total evidence is permissible if the balance of the evidence is irrelevant to the inductive "conclusion," i.e., if on the partial evidence alone, the conclusion has the same confirmation, or logical probability, as on the total evidence.¹

The requirement of total evidence is not a postulate nor a theorem of inductive logic; it is not concerned with the formal validity of inductive arguments. Rather, as Carnap has stressed, it is a maxim for the *application* of inductive logic; we might say that it states a necessary condition of

rationality of any such application in a given "knowledge situation," which we will think of as represented by the set K of all statements accepted in the situation.

But in what manner should the basic idea of this requirement be brought to bear upon probabilistic explanation? Surely we should not insist that the explanans must contain all and only the empirical information available at the time. Not all the available information, because otherwise all probabilistic explanations acceptable at time t would have to have the same explanans, K_t ; and not only the available information, because a proffered explanation may meet the intent of the requirement in not overlooking any relevant information available, and may nevertheless invoke some explanans statements which have not as yet been sufficiently tested to be included in K_t .

The extent to which the requirement of total evidence should be imposed upon statistical explanations is suggested by considerations such as the following. A proffered explanation of Jones's recovery based on the information that Jones had a streptococcal infection and was treated with penicillin, and that the statistical probability for recovery in such cases is very high, is unacceptable if K includes the further information that Jones's streptococci were resistant to penicillin, or that Jones was an octogenarian with a weak heart, and that in these reference classes the probability of recovery is small. Indeed, one would want an acceptable explanation to be based on a statistical probability statement pertaining to the narrowest reference class of which, according to our total information, the particular occurrence under consideration is a member. Thus, if K tells us not only that Jones had a streptococcus infection and was treated with penicillin, but also that he was an octogenarian with a weak heart (and if K provides no information more specific than that) then we would require that an acceptable explanation of Jones's response to the treatment be based on a statistical law stating the probability of that response in the narrowest reference class to which our total information assigns Jones's illness, i.e., the class of streptococcal infections suffered by octogenarians with weak hearts.⁹

Let me amplify this suggestion by reference to an example concerning the use of the law that the half-life of radon is 3.82 days in accounting for the fact that the residual amount of radon to which a sample of 10 milligrams was reduced in 7.64 days was within the range from 2.4 to 2.6 milligrams. According to present scientific knowledge, the rate of decay of a radioactive element depends solely upon its atomic structure as characterized by its atomic number and its mass number, and it is thus unaffected by the age of the sample and by such factors as temperature, pressure, magnetic and electric forces, and chemical interactions. Thus, by specifying the half-life of radon as well as the initial mass of the sample and the time interval in question, the explanans takes into account all the

available information that is relevant to appraising the probability of the given outcome by means of statistical laws. To state the point somewhat differently: Under the circumstances here assumed, our total information K assigns the case under study first of all to the reference class say F_1 , of cases where a 10 milligram sample of radon is allowed to decay for 7.64 days; and the half-life law for radon assigns a very high probability, within F_1 , to the "outcome," say G , consisting in the fact that the residual mass of radon lies between 2.4 and 2.6 milligrams. Suppose now that K also contains information about the temperature of the given sample, the pressure and relative humidity under which it is kept, the surrounding electric and magnetic conditions, and so forth, so that K assigns the given case to a reference class much narrower than F_1 , let us say, $F_1 F_2 F_3 \dots F_n$. Now the theory of radioactive decay, which is equally included in K , tells us that the statistical probability of G within this narrower class is the same as within G . For this reason, it suffices in our explanation to rely on the probability $p(G, F_1)$.

Let us note, however, that "knowledge situations" are conceivable in which the same argument would not be an acceptable explanation. Suppose, for example, that in the case of the radon sample under study, the amount remaining one hour before the end of the 7.64 day period happens to have been measured and found to be 2.7 milligrams, and thus markedly in excess of 2.6 milligrams—an occurrence which, considering the decay law for radon, is highly improbable, but not impossible. That finding, which then forms part of the total evidence K , assigns the particular case at hand to a reference class, say F^* , within which, according to the decay law for radon, the outcome G is highly improbable since it would require a quite unusual spurt in the decay of the given sample to reduce the 2.7 milligrams, within the one final hour of the test, to an amount falling between 2.4 and 2.6 milligrams. Hence, the additional information here considered may not be disregarded, and an explanation of the observed outcome will be acceptable only if it takes account of the probability of G in the narrower reference class, i.e., $p(G, F_1 F^*)$. (The theory of radioactive decay implies that this probability equals $p(G, F^*)$, so that as a consequence the membership of the given case in F_1 need not be explicitly taken into account.)

The requirement suggested by the preceding considerations can now be stated more explicitly; we will call it the *requirement of maximal specificity for inductive-statistical explanations*. Consider a proposed explanation of the basic statistical form

$$\begin{array}{l}
 3a \quad p(G, F) = r \\
 \quad \frac{Fb}{Gb} \quad [r]
 \end{array}$$

Let s be the conjunction of the premises, and, if K is the set of all statements accepted at the given time, let k be a sentence that is logically equivalent to K (in the sense that k is implied by K and in turn implies every sentence in K). Then, to be rationally acceptable in the knowledge situation represented by K , the proposed explanation (3a) must meet the following condition (the requirement of maximal specificity): If $s \cdot k$ implies⁶ that b belongs to a class F_1 , and that F_1 is a subclass of F , then $s \cdot k$ must also imply a statement specifying the statistical probability of G in F_1 , say

$$p(G, F_1) = r_1$$

Here, r_1 must equal r unless the probability statement just cited is simply a theorem of mathematical probability theory.

The qualifying unless-clause here appended is quite proper, and its omission would result in undesirable consequences. It is proper because theorems of pure mathematical probability theory cannot provide an explanation of empirical subject matter. They may therefore be discounted when we inquire whether $s \cdot k$ might not give us statistical laws specifying the probability of G in reference classes narrower than F . And the omission of the clause would prove troublesome, for if (3a) is proffered as an explanation, then it is presumably accepted as a fact that Gb ; hence ' Gb ' belongs to K . Thus K assigns b to the narrower class $F \cdot G$, and concerning the probability of G in that class, $s \cdot k$ trivially implies the statement that $p(G, F \cdot G) = 1$, which is simply a consequence of the measure-theoretical postulates for statistical probability. Since $s \cdot k$ thus implies a more specific probability statement for G than that invoked in (3a), the requirement of maximal specificity would be violated by (3a)—and analogously by any proffered statistical explanation of an event that we take to have occurred—were it not for the unless-clause, which, in effect, disqualifies the notion that the statement ' $p(G, F \cdot G) = 1$ ' affords a more appropriate law to account for the presumed fact that Gb .

The requirement of maximal specificity, then, is here tentatively put forward as characterizing the extent to which the requirement of total evidence properly applies to inductive-statistical explanations. The general idea thus suggested comes to this: In formulating or appraising an I-S explanation, we should take into account all that information provided by K which is of potential explanatory relevance to the explanandum event; i.e., all pertinent statistical laws, and such particular facts as might be connected, by the statistical laws, with the explanandum event.⁷

The requirement of maximal specificity disposes of the problem of epistemic ambiguity; for it is readily seen that of two rival statistical arguments with high associated probabilities and with premises that all belong to K , at least one violates the requirement of maximum specificity. Indeed, let

$$\frac{p(G, F) = r_1}{\frac{Fb}{Gb}} [r_1] \quad \text{and} \quad \frac{p(\bar{G}, H) = r_2}{\frac{Hb}{\bar{G}b}} [r_2]$$

be the arguments in question, with r_1 and r_2 close to 1. Then, since K contains the premises of both arguments, it assigns b to both F and H and hence to $F \cdot H$. Hence if both arguments satisfy the requirement of maximal specificity, K must imply that

$$\begin{aligned} p(G, F \cdot H) &= p(G, F) = r_1 \\ p(\bar{G}, F \cdot H) &= p(\bar{G}, H) = r_2 \\ \text{But } p(G, F \cdot H) + p(\bar{G}, F \cdot H) &= 1 \\ \text{Hence } r_1 + r_2 &= 1 \end{aligned}$$

and this is an arithmetic falsehood, since r_1 and r_2 are both close to 1; hence it cannot be implied by the consistent class K .

Thus, for I-S explanations that meet the requirement of maximal specificity the problem of epistemic ambiguity no longer arises. We are never in a position to say: No matter whether this particular event did or did not occur, we can produce an acceptable explanation of either outcome: and an explanation, moreover, whose premises are scientifically accepted statements which confer a high logical probability upon the given outcome.

While the problem of epistemic ambiguity has thus been resolved, ambiguity in the first sense discussed [in section 2] remains unaffected by our requirement; i.e., it remains the case that for a given statistical argument with true premises and a high associated probability, there may exist a rival one with equally true premises and with a high associated probability, whose conclusion contradicts that of the first argument. And though the set K of statements accepted at any time never includes all statements that are in fact true (and no doubt many that are false), it is perfectly possible that K should contain the premises of two such conflicting arguments; but as we have seen, at least one of the latter will fail to be rationally acceptable because it violates the requirement of maximal specificity.

The preceding considerations show that *the concept of statistical explanation for particular events is essentially relative to a given knowledge situation as represented by a class K of accepted statements*. Indeed, the requirement of maximal specificity makes explicit and unavoidable reference to such a class, and it thus serves to characterize the concept of "I-S explanation relative to the knowledge situation represented by K ." We will refer to this characteristic as the *epistemic relativity of statistical explanation*.

It might seem that the concept of deductive explanation possesses the

same kind of relativity, since whether a proposed D-N or D-S [deductive-statistical] account is acceptable will depend not only on whether it is deductively valid and makes essential use of the proper type of general law, but also on whether its premises are well supported by the relevant evidence at hand. Quite so; and this condition of empirical confirmation applies equally to statistical explanations that are to be acceptable in a given knowledge situation. But the epistemic relativity that the requirement of maximal specificity implies for I-S explanations is of quite a different kind and has no analogue for D-N explanations. For the specificity requirement is not concerned with the evidential support that the total evidence *K* affords for the explanans statements: it does not demand that the latter be included in *K*, nor even that *K* supply supporting evidence for them. It rather concerns what may be called the concept of a *potential* statistical explanation. For it stipulates that no matter how much evidential support there may be for the explanans, a proposed I-S explanation is not acceptable if its potential explanatory force with respect to the specified explanandum is vitiated by statistical laws which are included in *K* but not in the explanans, and which might permit the production of rival statistical arguments. As we have seen, this danger never arises for deductive explanations. Hence, these are not subject to any such restrictive condition, and the notion of a potential deductive explanation (as contrasted with a deductive explanation with well-confirmed explanans) requires no relativization with respect to *K*.

As a consequence, we can significantly speak of true D-N and D-S explanations: they are those potential D-N and D-S explanations whose premises (and hence also conclusions) are true—no matter whether this happens to be known or believed, and thus no matter whether the premises are included in *K*. But this idea has no significant analogue for I-S explanation since, as we have seen, the concept of potential statistical explanation requires relativization with respect to *K*.

■ | Notes

1. Phrases such as 'It is almost certain (very likely) that *j* recovers', even when given the relational construal here suggested, are ostensibly concerned with relations between propositions, such as those expressed by the sentences forming the conclusion and the premises of an argument. For the purpose of the present discussion, however, involvement with propositions can be avoided by construing the phrases in question as expressing logical relations between corresponding sentences, e.g., the conclusion-sentence and the premise-sentence of an argument. This construal, which underlies the formulation of (1c), will be adopted in this essay, though for the sake of convenience we may occasionally use a paraphrase.

2. In the familiar schematization of deductive arguments, with a single line separating the premises from the conclusion, no explicit distinction is made between a weaker and a stronger claim, either of which might be intended; namely (i) that the premises logically imply the conclusion and (ii) that, in addition, the premises are true. In the case of our probabilistic argument, (1c) expresses a weaker claim, analogous to (i), whereas (1d) may be taken to express a "proffered explanation" (the term is borrowed from I. Scheffler, 'Explanation, Prediction, and Abstraction', *British Journal for the Philosophy of Science* 7 (1957), sect. 1) in which, in addition, the explanatory premises are—however tentatively—asserted as true.

The considerations here outlined concerning the use of terms like 'probably' and 'certainly' as modal qualifiers of individual statements seem to me to militate also against the notion of categorical probability statement that C. I. Lewis sets forth in the following passage (italics the author's):

Just as 'If D then (certainly) P, and D is the fact', leads to the categorical consequence, 'Therefore (certainly) P'; so too, 'If D then probably P, and D is the fact', leads to a categorical consequence expressed by 'It is probable that P'. And this conclusion is not merely the statement over again of the probability relation between 'P' and 'D'; any more than 'Therefore (certainly) P' is the statement over again of 'If D then (certainly) P'. If the barometer is high, tomorrow will probably be fair; and the barometer is high', categorically assures something expressed by 'Tomorrow will probably be fair'. This probability is still relative to the grounds of judgment; but if these grounds are actual, and contain all the available evidence which is pertinent, then it is not only categorical but may fairly be called the probability of the event in question (1946: 319).

This position seems to me to be open to just those objections suggested in the main text. If 'P' is a statement, then the expressions 'certainly P' and 'probably P' as envisaged in the quoted passage are not statements. If we ask how one would go about trying to ascertain whether they were true, we realize that we are entirely at a loss unless and until a reference set of statements or assumptions has been specified relative to which P may then be found to be certain, or to be highly probable, or neither. The expressions in question, then, are essentially incomplete; they are elliptic formulations of relational statements; neither of them can be the conclusion of an inference. However plausible Lewis's suggestion may seem, there is no analogue in inductive logic to *modus ponens*, or the "rule of detachment," of deductive logic, which, given the information that 'D' and also 'if D then P', are true statements, authorizes us to detach the consequent 'P' in the conditional premise and to assert it as a self-contained statement which must then be true as well.

At the end of the quoted passage, Lewis suggests the important idea that 'probably P' might be taken to mean that the total relevant evidence available at the time confers high probability upon P. But even this statement is relational in that it tacitly refers to some unspecified time, and, besides, his general notion of a categorical probability statement as a conclusion of an argument is not made dependent on the assumption that the premises of the argument include all the relevant evidence available.

It must be stressed, however, that elsewhere in his discussion, Lewis emphasizes the relativity of (logical) probability, and, thus, the very characteristic that rules out the conception of categorical probability statements.

Similar objections apply, I think, to Toulmin's construal of probabilistic arguments; cf. Toulmin (1958) and the discussion in Hempel (1960), sects. 1-3.

3. R. Carnap, *Logical Foundations of Probability* (Chicago, 1950), 211. The requirement is suggested, e.g., in the passage from Lewis quoted in n. [2]. Similarly Williams speaks of "the most fundamental of all rules of probability logic, that 'the' probability of any proposition is its probability in relation to the known premises and them only" (*The Ground of Induction* (Cambridge, Mass., 1947), 72).

I am greatly indebted to Professor Carnap for having pointed out to me in 1945, when I first noticed the ambiguity of probabilistic arguments, that this was but one of several apparent paradoxes of inductive logic that result from disregard of the requirement of total evidence.

S. F. Barker, *Induction and Hypothesis* (Ithaca, NY, 1957), 70-78, has given a lucid independent presentation of the basic ambiguity of probabilistic arguments, and a skeptical appraisal of the requirement of total evidence as a means of dealing with the problem. However, I will presently suggest a way of remedying the ambiguity of probabilistic explanation with the help of a rather severely modified version of the requirement of total evidence. It will be called the requirement of maximal specificity, and is not open to the same criticism.

4. Cf. Carnap, *Logical Foundations*, 211 and 494.

5. This idea is closely related to one used by H. Reichenbach, (cf. *The Theory of Probability* (Berkeley, Calif., and Los Angeles, 1949), sect. 72) in an attempt to show that it is possible to assign probabilities to individual events within the framework of a strictly statistical conception of probability. Reichenbach proposed that the probability of a single event, such as the safe completion of a particular scheduled flight of a given commercial plane, be construed as the statistical probability which the *kind* of event considered (safe completion of a flight) possesses within the narrowest reference class to which the given case (the specified flight of the given plane) belongs, and for which reliable statistical information is available (e.g., the class of scheduled flights undertaken so far by planes of the line to which the given plane belongs, and under weather conditions similar to those prevailing at the time of the flight in question).

6. Reference to $s \cdot k$ rather than to k is called for because, as was noted earlier, we do not construe the condition here under discussion as requiring that all the explanans statements invoked be scientifically accepted at the time in question, and thus be included in the corresponding class K .

7. By its reliance on this general idea, and specifically on the requirement of maximal specificity, the method here suggested for eliminating the epistemic ambiguity of statistical explanation differs substantially from the way in which I attempted in an earlier study (Hempel, 'Deductive-Nomological vs. Statistical Explanation', esp. sect. 10) to deal with the same problem. In that study, which did not distinguish explicitly between the two types of explanatory ambiguity characterized earlier in this section, I applied the requirement of total evidence to statistical explanations in a manner which presupposed that the explanans of any acceptable explanation belongs to the class K , and which then demanded that the probability which the explanans confers upon the explanandum be equal to that which the total evidence, K , imparts to the explanandum. The reasons why this approach seems unsatisfactory to me are suggested by the arguments set forth in the present section. Note in particular that, if strictly enforced, the requirement of total evidence would preclude the possibility of any significant statistical expla-

nation for events whose occurrence is regarded as an established fact in science: for any sentence describing such an occurrence is logically implied by *K* and thus trivially has the logical probability 1 relative to *K*.

■ | References

Barker, S. F., *Induction and Hypothesis* (Ithaca, NY: Cornell University Press, 1957).

Carnap, R., *Logical Foundations of Probability* (Chicago: University of Chicago Press 1950; second, rev., edn. 1962).

Feigl, H., and Maxwell, G. (eds.), *Minnesota Studies in the Philosophy of Science*, iii (Minneapolis: University of Minnesota Press, 1962).

Hempel, C. G., 'Inductive Inconsistencies', in *Synthese*, 12 (1960), 439-69, repr. in *Aspects of Scientific Explanation* (New York and London: Free Press and Collier Macmillan, 1965), 53-79.

——— 'Deductive-Nomological vs. Statistical Explanation', in Feigl and Maxwell (eds.), 98-169.

Lewis, C. I., *An Analysis of Knowledge and Valuation* (La Salle, Ill.: Open Court, 1946).

Reichenbach, H., *The Theory of Probability* (Berkeley, Calif., and Los Angeles: The University of California Press, 1949).

Scheffler, I., 'Explanation, Prediction, and Abstraction', *The British Journal for the Philosophy of Science* 7 (1957) 293-309.

Toulmin, S., *The Uses of Argument* (Cambridge: Cambridge University Press, 1958).

Williams, D. C., *The Ground of Induction* (Cambridge, Mass.: Harvard University Press, 1947).

Arguments, Laws, and Explanation

I start by way of outlining some of the standard counterexamples to Hempel's account of full explanation. . . .

These counterexamples cluster around two difficulties: (A) irrelevance and (B) symmetry. I do not say that Hempel's account has no resources for replying adequately to *any* of these standard counterexamples, although I do think that this is true in some cases. I indicate where I believe that this is so. The counterexamples purport to show that Hempel's account of explanation, even if necessary, could not be sufficient. However, I argue that in thinking through an adequate response to the counterexamples, we will see that Hempel's requirements are not even necessary for (full) explanation. . . .

■ | The Standard Counterexamples: Irrelevance

The first reason, (A), for holding that Hempel's conditions for explanation could not be sufficient turns on the fact that there can be derivations that meet all of Hempel's requirements for D-N (or I-S) explanation, but whose premises are obviously irrelevant to the explanation of the conclusions of those derivations. In the main, I shall only be concerned in this chapter with the explanation of singular facts . . . , but we might note some counterexamples which concern the explanation of laws as well. Here is one, taken from Ardon Lyon, which concerns the explanation of empirical laws by deductive subsumption.¹

1 All metals conduct electricity.

2 Whatever conducts electricity is subject to gravitational attraction.

∴ 3 All metals are subject to gravitational attraction.

As Lyon points out, no one would regard the conjunction of (1) and (2) as explaining (3), in spite of the fact that the latter does follow from the former, because (1) and (2) are irrelevant to the truth of (3). 'Metals are not subject to gravitational attraction because they conduct electricity: non-conductors are subject to gravitational attraction to just the same degree' (Lyon 1974: 247). Lyon's counterexample is directed against Hempel's account of the explanation of laws, but it is easy to construct a parallel counterexample to Hempel's account of the explanation of singular facts. The explanandum in question would be that this bit of metal is subject to gravitational attraction, and the explanans will include the fact that this bit of metal conducts electricity.

Another alleged counterexample to Hempel's analysis of the explanation of laws is offered by Baruch Brody.²

- 1 Sodium normally combines with bromine in a ratio of one-to-one.
- 2 Everything that normally combines with bromine in a ratio of one-to-one normally combines with chlorine in a ratio of one-to-one.

- ∴ 3 Sodium normally combines with chlorine in a ratio of one-to-one.

Brody claims that this derivation has no explanatory power whatever, and I agree with him. But even if the reader were to insist that it has some such power, it doesn't have much, and Hempel's analysis does not offer us the materials for saying why that should be so. Although Brody does not say so, one could say that the problem here too is one of explanatory irrelevance. The ratio in which bromine and chlorine combine is surely irrelevant for explaining (but not necessarily irrelevant in other ways) the ratio in which sodium and chlorine combine, even though the two ratios are related in a lawlike manner. As with Lyon's counterexample, it is simple to convert Brody's counterexample to one concerning the explanation of a singular fact: the fact that this bit of sodium combined with this bit of chlorine in a one-one ratio.

Two further counterexamples which I wish to mention are specifically directed to irrelevance in the case of the explanation of singular facts. The first example is adapted from Peter Achinstein.³ Suppose that poor Jones (he is so often ill) eats at least a pound of arsenic and dies within twenty-four hours, and that eating at least a pound of arsenic inevitably leads to death within twenty-four hours. Does it follow that the argument below is an explanation of Jones's death?

- 1 Jones ate at least a pound of arsenic at time t .
 - 2 (x) (x eats at least 1 lb arsenic at t \supset x dies within 24 hours after t).
-
- \therefore 3 Jones dies within 24 hours of t .

Suppose, consistently with the above suppositions, that Jones was run over by a bus and died soon after ingesting the arsenic. In this case, the deduction will not be explanatory, since Jones, although he would have died from the arsenic had he not been run over by a bus soon after eating the poison, was actually killed by the bus. It is the bus, and not the arsenic, which explains his death, in spite of the argument given above meeting all of Hempel's conditions.

One can generalize Achinstein's example, to any case in which there is causal pre-emption. Suppose some event, e , has two potential causes c and d , in the sense that c occurs and causes e , and that d also occurs and does not cause e , but would have caused e if c had not occurred. d is a potential alternative cause of e , but is pre-empted by the actual cause c .⁴ In any such case, there will be an Achinstein-style counterexample to the D-N account of the explanation of singular facts, since there will be a derivation (with all true premises, etc.) to the explanandum via a premiss set which includes a premiss about the pre-empted cause but not one about the actual cause, and hence no explanation of the explanandum so derived. The pre-empted cause is explanatorily irrelevant to the explanandum thus derived.

I do take the lesson of this counterexample to be important, so it will be worth dwelling on it. Is there a way of meeting this alleged counterexample from the existing resources of Hempel's theory? One might think that it can be met by the introduction of a *ceteris paribus* ["other things being equal"] clause in the statement of the law, (2), and the addition of a further premiss (which will in this case be false) that says that other conditions are in fact equal.⁵ So the 'irrelevant explanation', since it includes a false premiss, will fail to be an explanation on Hempel's own account. After all, the rejoinder goes, no one can die who is already dead; the arsenic will be what kills Jones only if he hasn't already died from some other cause. The arsenic ingestion is relevant, only if the *ceteris paribus* clause in the law is met, and the clause will exclude the case in which an alternative cause operates.

I fail to see how the *ceteris paribus* clause response will meet the difficulty at hand. A *ceteris paribus* clause is inserted in a law, as a means of saving an apparently falsified law from real falsification: other things are not equal, so the law is true after all. However, in the counterexample, Jones's being run over by a bus does not even apparently falsify the law that whoever eats at least a pound of arsenic dies within twenty-four hours. After all, after eating the arsenic, Jones *did die* within the required time

period. So how could Jones's bus-related death present any kind of even apparent difficulty for the law about what happens to people after they ingest at least a pound of arsenic? Any difficulty for that law must involve someone's failure to die in some circumstances or other, and poor, dead Jones is no example of that.

In general, when *c* (the bus hitting Jones) causes *e* (the death of Jones), there is no argument from this fact to the falsity of the law that whenever a *D* (an ingestion of at least a pound of arsenic), then an *E* (a death). In particular, one does not need to rephrase the law as: Whenever a *D*, then an *E*, unless there is some alternative cause that operates to bring about an *E*. It is *true* that whoever eats a pound of arsenic at *t* dies within twenty-four hours, even when sometimes death of arsenic ingestors is actually brought about by buses or something else.

One further reply to this counterexample might dispute that (2) correctly expresses the intended law. Suppose we interpret the law as itself including a causal claim: eating a pound of arsenic *causes* death within twenty-four hours. If laws are universally quantified generalizations (remember that we are assuming throughout that this is so), how should we represent 'eating 1 lb arsenic at *t* causes death within 24 hours', in such a way that it would retain an explicit causal claim? Perhaps in this way: (x) (x eats at least 1 lb arsenic at *t* \supset x's eating at least 1 lb arsenic at *t* causes x's death within 24 hours). There may be some other way in which to capture the causal claim in an explicit way within the universally quantified generalization, but I cannot see what it might be.

This generalization is falsified by the case in which Jones eats the arsenic but the bus causes his death, so a *ceteris paribus* clause would have to be inserted into it after all. If this is the law, it surely intends to assert that one's eating that much arsenic will cause death, *unless* something else causes it instead. The qualification, 'unless something else causes it instead', would be included in the *ceteris paribus* clause. The law should therefore be expressed as: (2') (x) (x eats at least 1 lb arsenic at *t* & *ceteris paribus* \supset x's eating at least 1 lb arsenic at *t* causes x's death within 24 hours). The explanatory argument which uses (2') would have to include an additional premise: (2'') Other things are equal. If the bus and not the arsenic kills Jones, (2'') would be false, and so the argument would fail to be explanatory, on Hempel's own account. Can we conclude then that, on this view of what the law is, the *ceteris paribus* strategy could handle the arsenic-and-bus counterexample to Hempel's account after all?

I think not, for two reasons. First, this strategy is simply not available to Hempel. No supporter (like Hempel) of the orthodox view of laws would accept (2') as giving the correct form for a causal law. Second, there are additional problems about what the explanandum would be which (2') would help to explain; the explanandum certainly would not be as given by (3), 'Jones dies within 24 hours of *t*'. The explanandum explained by (2') could only be: (3') 'eating at least 1 lb arsenic caused

Jones's death within 24 hours of t' . One might wrongly suppose that this will present no difficulty for Hempel, since (3) follows from (3'). If one explains (3'), and (3) follows from (3'), hasn't one explained (3) as well?

As Peter Lipton has pointed out,⁶ this assumption is not available to Hempel. Hempel's D-N model of explanation is itself not closed under logical entailment. Suppose conclusion c is derived from, and explained by, law L and initial conditions i . The disjunction, i or c , logically follows from c . But the explanation of c by the conjunction of L and i cannot, on Hempel's account, be an explanation of i or c , because L is not essential to the derivation of i or c from the conjunction of L and i .

I have no doubt there is some way to handle the arsenic-and-bus case, but the introduction of a *ceteris paribus* clause into the law is simply not it. Nor do I think that there are any resources available in Hempel's account as it stands for satisfactorily dealing with it.

The arsenic-and-bus counterexample is interesting for another reason. It provides an additional example of the asymmetry between explanation and prediction. Someone who produces the above argument, (1)–(3), cannot be said to have explained Jones's death, but he certainly will have been able to predict it successfully. He predicts that Jones will die, and his prediction is correct. Moreover, he has offered excellent grounds for his prediction. Given that Jones drank the arsenic, the predictor could be certain that Jones would die. One can predict via a pre-empted cause, even though one cannot explain via one. Any rejoinder which wishes to claim that the above argument yields neither a successful prediction nor a successful explanation will owe us a fuller account of successful prediction than has been, thus far at any rate, provided.

A second example of explanatory irrelevance which constitutes a counterexample to Hempel's analysis of explanation of singular facts is one taken from Wesley Salmon.⁷

- 1 Every man who regularly takes birth control pills avoids pregnancy.
- 2 John Jones has taken his wife's birth control pills regularly.
- ∴ 3 John Jones avoided becoming pregnant in the past year.

The same sort of case can be made out for someone 'who explains the dissolving of a piece of sugar by citing the fact that the liquid in which it dissolved is holy water'. A sentence which states the fact that the sugar dissolved in that liquid can be derived from, but hardly explained by, sentences stating the fact that the liquid is holy water and the relevant law connecting water and the dissolution of sugar. The fact that the water is holy water is not relevant to the explanation of the dissolution. If (1) above is rephrased as a stochastic rather than as a deterministic law, it will serve

as an irrelevance counterexample to the Hempelian account of I-S explanation.

A determined advocate of Hempel's models of explanation might try to insist that the inclusion of explanatorily irrelevant material in the explanans in Salmon's counterexamples might make the explanations poor(er), but that they are still explanations none the less. [Elsewhere] I distinguished between cases in which an explanation is bad and cases in which there is no explanation at all, and . . . I applied that distinction specifically to the example of the inclusion of explanatorily irrelevant information in the explanans. . . .^{*} [For example] that someone is a man who takes birth control pills entails that the person is a man, and the person's being a man explains why that person does not become pregnant, but the person's being a man who takes birth control pills does not explain in the least why the person does not become pregnant.

I agree with Salmon about this. . . . The richer information has the explanatorily relevant information buried in it: its being water is included in its being holy water; the person's being a man is included in the person's being a man who takes birth control pills. But the richer information does not explain some explanandum, just in virtue of the fact that the weaker information which it includes and hence entails does explain it. The additional information which makes it richer but which is explanatorily irrelevant overrides and kills the explanatory power of the weaker information when it is added to it. As Salmon said, irrelevance is *fatal* to explanation.

The examples which I group under (A) all teach the same lesson. There can be derivations which meet all of Hempel's conditions for the explanation of a singular fact, but, whereas they are wonderful derivations, they offer no explanation of what is derived. This is because the premisses are explanatorily irrelevant to the conclusion, or contain misleading explanatorily irrelevant additional information, even though they do imply the conclusion. . . .

■ | The Standard Counterexamples: Symmetry

The second reason, (B), for holding that Hempel's conditions could not be sufficient for singular explanation has to do with 'explanatory' symmetries. Hempel's account of singular explanation in terms of derivability from true, empirical premisses permits intuitively objectionable cases in

^{*} Ruben argues for the bad explanation—no explanation distinction and the thesis that adding irrelevant information to the premisses of an argument can rob it of all explanatory power in chapters 1 and 5, respectively, of his *Explaining Explanation*, from which the present reading is taken.

which (part of) the explanans can be explained by the explanandum, as well as explain it. How can we amend the account, so that such symmetries of 'explanation' will not arise?

Both James Woodward and Peter Achinstein have argued (or implied) that the explanation relation is not an asymmetric relation, as is usually supposed, and that there are or can be bona fide cases of acceptable symmetrical explanation, explanatory mutual dependence between two singular facts.⁸ However, both would of course concede that there are some cases in which symmetrical explanation must be ruled out (i.e. in the case of causal explanation). The explanation relation, even if not asymmetric, is surely not symmetric. If not asymmetric, it must be non-symmetric. This is enough for my argument here. All the examples I shall consider in this part of the chapter are cases in which symmetrical explanations are intuitively unacceptable; I do not need to retain the stronger claim that the explanation relation itself is asymmetric. . . .

There are a number of these 'symmetry' counterexamples which challenge Hempel's account of singular explanation, many of which derive from Sylvain Bromberger and Michael Scriven.⁹ We have already touched on some of these examples in the discussion of Hempel. There are really two kinds of cases that generate these unacceptable symmetries. First, there are equations which show that the numerical value assumed by some property of a system at time t is a function of the values assumed by other properties of a system at time t or an earlier time, $t - \Delta$ (Ohm's law, Hooke's law, the Boyle-Charles laws for ideal gases, the length and period of a pendulum).

Second, there are laws with biconditionals, which can include cases both of laws of coexistence and of laws of succession. A barometer falls iff [if and only if] a storm is approaching; the light received from the galaxies exhibits a shift towards the red end of the spectrum iff the galaxies are receding from us; and (Aristotle's case) a planet twinkles iff it is not near. To this, we can add Salmon's confused rooster who explains the rising of the sun on the grounds of his regular crowing.¹⁰ These equations or biconditionals will allow the derivation of the height of the flagpole from the length of the shadow and the length of the shadow from the height of the flagpole; the length of the pendulum from its period and its period from its length; the approaching storm from the fall in the barometer as well as the fall of the barometer from the approaching storm; the receding of the galaxies from the red shift as well as the red shift from the recession of the galaxies, the rising of the sun from the crowing of the cock as well as the crowing of the cock from the rising of the sun.

But, in each of these pairs, the first derivation would be nonexplanatory; the second, explanatory. Equations and biconditionals permit symmetric derivations; but since at least these examples do not provide symmetric explanations, there must be more to singular explanation than what Hempel's theory thus far allows.

Hempel, as we saw, 'dealt' with this by suggesting that there may not really be true biconditionals in such cases (he supposed, it will be recalled, that there might be cases of Koplik spots without measles). But what we have to establish is how, given that there may really be true biconditionals or equations of this kind which allow derivations 'in both directions', we are able to distinguish the explanations from the derivations which fail to explain.

■ | A Proposed Cure and Its Problems: The Causal Condition

It is not a novel thought that the cure for the problems of irrelevance and symmetry, (A) and (B), that Hempel's analysis of D-N explanation faces (at least for the explanation of singular facts; explanation of laws would be quite a different matter) is to be found by stipulating that the premisses include something about the cause of the event to be explained. This was Aristotle's suggestion . . . for the examples of the [non]twinkling planets and the deciduous vines.* Mill's official theory, which requires that the premisses include the statement of a causal law, has similar resources for dealing with the counterexamples. At least some explanations are, on such an account, deductively valid arguments with true premisses which have empirical content, one of which is a lawlike generalization (thus far, Aristotle, Mill, and Hempel can agree), but also one of which mentions or specifies in some way the cause of the explanandum event (the final requirement would have to be added to the Hempelian account, but is already explicit in the accounts of the other two).

* Aristotle's two examples of the nontwinkling planets and the deciduous vines can be found in his *Posterior Analytics*, book I, chapter 13, and book II, chapter 16, respectively. Both illustrate Aristotle's distinction between *knowledge of the fact* and *knowledge of the reasoned fact*. Knowledge of the fact occurs when we use a deductive argument with true premisses to infer that something is the case. Knowledge of the reasoned fact—*explanatory knowledge*—occurs only when we reason (deductively) from causes to effects, thus yielding knowledge of *why* something is the case. For example, Aristotle believed (falsely, as it turns out) that all and only broad-leaved plants are deciduous. Thus, one could correctly infer that a vine is broad-leaved from the fact that it is deciduous. But knowledge of the reasoned fact occurs only when we infer that a vine is deciduous from the fact that it is broad-leaved. For only in that case, according to Aristotle, would we understand *why* the vine is deciduous. (Aristotle assumes that having broad leaves is what causes a plant to shed its leaves but that shedding of leaves is not what causes a plant to have broad leaves.) In a similar way, Aristotle attributes knowledge of the fact to someone who infers that planets are near the earth from the fact that they do not twinkle, but he attributes knowledge of the reasoned fact to someone who infers that planets do not twinkle from the fact that they are near the earth. (Proximity to the earth causes celestial objects to shine steadily, but their shining steadily does not cause them to be near the earth.)

How would the causal requirement help with the problem of symmetry? Given the angle of the sun's elevation, it is the height of the flagpole that causes the length of the shadow, and not vice versa; the change in atmospheric pressure that causes the rise or fall of the barometer, and not vice versa; the receding of the galaxies that causes the red shift, and not vice versa. The causal requirement will also help with irrelevance. It was the bus but not the arsenic, his being a man but not his taking birth control pills, the substance's being water but not its being holy water, which is causally relevant to the death of Jones, the pregnancy failure, and the dissolution of the sugar. So causation seems a way both to rule out symmetric 'explanations' (anyway, where these are unwelcome) and irrelevant 'explanations'.

One might doubt whether causation will in fact help with irrelevance. Suppose we have a jar in which there is some sugar. We add to the sugar some water appropriately blessed by the local priest. What caused the dissolution of the sugar? In part, its being immersed in the water. But, the sample of water just is a sample of holy water, so if the immersion in the water caused the dissolution, then the immersion in the holy water caused it. If the immersion in the water not only caused but also explains the sugar's dissolution, doesn't the immersion in the holy water explain it too?

No, for we have distinguished . . . between causation and causal explanation. It is true that the immersion in the water, and hence in the holy water, causes the dissolution of the sugar. But it is the *fact* that it was immersed in the water in the jar that *causally explains* the fact that it dissolved, and even though the water is holy water, the fact that it was immersed in the holy water in the jar is a different fact from the fact that it was immersed in the water in the jar. The fact that it was immersed in the holy water in the jar introduces a feature that the other fact does not introduce. And that additional feature, the water's being holy, is causally irrelevant to the dissolution and hence irrelevant to the explanation of the dissolution. A similar diagnosis will be available in the other cases of explanatory irrelevance we have looked at. The purported explaining facts introduce features which are causally irrelevant to what is being explained.

Many contemporary writers have converged on the necessity of including such a causal requirement. Thus, Salmon, reversing his earlier attempts to explicate explanation on the basis of statistical relations and without mention of causation, says that 'The explanatory significance of statistical relations is indirect. Their fundamental import lies in the fact . . . that they constitute evidence for causal relations', and 'The time has come . . . to put the "cause" back into "because"'.¹¹ Or, 'To give scientific explanations is to show how events and statistical regularities fit into the causal network of the world' (Salmon 1977: 162). Others, including Baruch Brody, have hit upon the same idea, of supplementing Hempel's account with some sort of causal information.¹²

The difficulty with this otherwise extremely attractive view has been pointed out by Timothy McCarthy.¹³ It is easy to construct examples of derivations which meet all of Hempel's conditions, plus the condition that there be a premiss which mentions the actual cause of the event to be explained, but which still fail to be explanatory. McCarthy has given several such examples.

His first example (slightly amended) is this. Let e be any event; let ' $D(e)$ ' represent any sentence describing e , and let ' $C(e)$ ' be a sentence which describes c , e 's actual (and not its pre-empted potential) cause (c is described under its causally relevant description). Let ' $(x)(Ax \supset Bx)$ ' represent any law utterly irrelevant to the occurrence of e . (It won't matter if you want to strengthen the requirement and make the law a causal law). Finally, let o be any object such that Ao . Consider the following derivation:

- 1 $(x)(Ax \supset Bx)$
- 2 $C(e) \ \& \ Ao$
- 3 $\sim B(o) \vee \sim C(e) \vee D(e)$
- \therefore 4 $D(e)$

This derivation of ' $D(e)$ ' from premisses (1)–(3) meets all of Hempel's conditions + the suggested causal supplement. e 's cause is described by ' $C(e)$ ' in premiss (2). Moreover, ' $C(e)$ ' is essential to the derivation (as is the law). Yet, no one would say that we have here an explanation of e , because even though c , e 's cause, is described in or mentioned by a premiss, it is not made causally and hence explanatorily relevant to e 's occurrence. There is still a notion of 'explanatory relevance' that 'derivation + mention of cause of what is to be explained' simply isn't getting at. As McCarthy says,

One might suppose that the idea is to mirror the causal dependence of e on its cause by the deductive dependence in d [the derivation] of a description of e upon a description of e 's cause. That is an interesting idea; immediately, however, we may begin to suspect a gap in the argument. The basic worry may be put in this way: why should it follow, merely because a D-N derivation of a sentence describing e ineliminably involves, in some way or other a description of e 's cause that this description functions in the derivation to show (causally) why e occurs? No obvious reason exists why a D-N derivation of a sentence describing e could not depend on a description of e 's cause in some way quite unrelated to the causal dependence of e on that cause. (McCarthy 1977: 161)

McCarthy shows that various attempts to outmanoeuvre this objection will fail. In particular, his argument can be sustained even if an additional

condition due to Kim is imposed.¹⁴ That condition is this: let all the singular sentences in the premisses be put in complete conjunctive normal form.* Then the condition requires that none of those conjuncts is a logical consequence of the explanandum itself. However, the following derivation meets all of Hempel's requirements + the causal requirement + Kim's conjunctive normal form condition. In the derivation below, 'C(o)' describes the cause of o's turning black, which, let us suppose, is o's being immersed in a bucket of black paint.

- 1 All crows are black.
 - 2 (x) (y) (x turns the colour of y & y is black \supset x turns black).
 - 3 C(o) & Henry is a crow.
 - 4 $\neg C(o) \vee o$ turns the colour of Henry.
-
- \therefore 5 o turns black.

Even though (1)–(4) meet all of Hempel's requirements + the causal supplement + Kim's condition, no explanation of (5) has been given.

There is surely something right in the demand that 'cause' be put back into 'because'. But what has gone wrong in the above examples? To simplify, in both derivations, call the cause 'c' and the effect to be explained, 'e'. Although it is true that one of the premisses in both of the above derivations says that c occurs, and although it is true that this premiss is essential to the derivation, no premiss asserts, *of c*, that it is the cause of e. The derivation gets us, as it were, to e's occurrence from c's occurrence, *not* via the fact that c causes e, but rather via a law irrelevant to c's causing e. There is no connection between c and e other than that of logical derivability of the latter's description from the former's (plus an irrelevant law), and that type of connection simply isn't enough to ensure explanation of the conclusion by the premisses. As McCarthy puts it:

The reason is precisely that the logical dependence of 'D(e)' on 'C(e)' has nothing at all to do with the causal dependence [and hence the explanatory dependence] of e on the event described by 'C(e)', because the law mediating the deductive relation between 'C(e)' and 'D(e)' is causally irrelevant to the occurrence of e.¹⁵

* Any truth-functional expression can be reduced to an equivalent expression in conjunctive normal form—that is, to a conjunction of disjunctions where every component of each disjunction is either an atomic proposition or the negation of one. Displaying all the nonlawlike (singular) premisses in this expanded form makes clear whether any part of the content of the explanans is entailed by the explanandum—this being the kind of circularity that Kim's constraint is designed to rule out.

In the note to the preceding sentence, I argue that various further attempts to strengthen the causal requirement, which require that the law not be irrelevant to the occurrence of the effect, will still leave us with non-explanatory derivations.

There is a very simple way to bring the cause and the explanandum event together in the right and relevant way, in order to ensure explanation: not by including as a premiss a singular statement which merely describes or mentions the cause of the explanandum event, *e*, but rather by including as a premiss a singular statement which asserts, of that cause, that it is the cause of *e*. The relevant premiss in McCarthy's arguments would say, for example, not only that *c* occurs, but also that *c* is the cause of *e*.¹⁶ If this were added, it seems that the derivation would become explanatory. And surely it is this that is lacking in McCarthy's examples, which accounts for the fact that they are not explanatory. This simple and expedient method avoids all the difficulties we have found in trying to capture explanatory dependence or relevance by logical dependence of conclusion on premisses. Explanatory dependence, at least in this example, is captured by an explicit statement of the causal dependence of the effect on the cause. Why just mention the cause in one of the premisses? Why shouldn't a premiss actually assert the causal dependence of explanandum event on explanans event?

I do not think that every such additional premiss must use the word 'cause'. The premiss might assert that *e* occurs *because* *c* occurs, or that the reason for *e* is *c*, or some such.¹⁷ In so far as we are here restricting ourselves to singular causal explanation, all of these will be ways of saying roughly the same thing. The point is this: the premiss under consideration will have to itself assert the dependence of effect on cause, and this dependence cannot be captured by logical dependence. The occurrence of the expression, '*c* is the cause of *e*', although frequently the way in which this is done, is hardly essential (remember that throughout I assume that the descriptions in the causal claim are the ones relevant for explanation); other alternative expressions, like ones which use 'because' or 'is the reason for', and which also capture this sense of non-logical dependence, will do equally well.

However, there are at least two important consequences of this last suggestion that we must note. First, *Hempel's* (and *Mill's*) requirement that there be a lawlike generalization in the premisses which is essential for the derivation is rendered unnecessary. On the suggestion being canvassed, we have in the argument a premiss that explicitly says: the cause of the explanandum event is such-and-such, and that premiss by itself will entail the statement that the explanandum event occurred, without the addition of any further premisses at all. In particular, no premiss stating a universal general fact, no law, will be required for the derivation of the explanandum. So the first consequence is the redundancy of laws in (at least some) explanations.

There is a second important consequence of this suggestion. Why think of explanations as arguments at all? True, we could think of the explanation as an argument with a single premiss:

- 1 c is the cause of e.
 ∴ 2 e.

But the derivation of 'e' from 'c is the cause of e' is trivial. It is simpler, and nothing is lost, if we think of this explanation as composed of a singular sentence, 'c is the cause of e' (or, 'e because of c', etc.). Since in fact all of the premisses save this one will be redundant, the explanation really just consists in the one remaining sentence that says that the cause of the event to be explained was such-and-such.

Deductivism and probabilism agreed that *all* full explanations are arguments; if McCarthy's argument and my elaboration of it above are sound, then at least sometimes full explanations are not arguments, but sentences. McCarthy's argument, in conjunction with my suggestion for remedying the defect to which it points, does not show that full explanations are never arguments; that conclusion would be too strong. But I would go further; typically, full explanations are not arguments, but singular sentences, or conjunctions thereof.

Is construing a specific bit of discourse as a sentence rather than an argument simply a matter of personal aesthetic preference on my part? McCarthy's argument and my subsequent remarks were intended to motivate the choice of sentence over argument. The explanation must explicitly include some word like 'because', 'reason', 'causes', etc., and it is just this that the idea of an explanatory argument was meant to avoid, by attempting to capture the dependence which such expressions get at by the idea of deductive or inductive logical dependence of a conclusion on premisses. We have seen how this strategy fails, and have seen that only explicit assertions or statements of the relevant dependence will do. Hence, such explanations typically consist, on my view, of sentences rather than arguments.

Let me mention one not very promising line of reply to this. Is there any *real* difference between an argument theory and a non-argument (or, specifically, a sentence) theory? Isn't the difference between an argument and a sentence theory somewhat superficial? There is, indeed, a way to trivialize the distinction between an argument and a sentence. Any argument can be rewritten as a conditional sentence, with the premisses as the antecedent and the conclusion as consequent. Such a conditional sentence, if true, is necessarily true. The explanatory sentences envisaged by a non-argument theory, if true, are contingently true. Explanations are typically contingently true sentences or conjunctions thereof. The sen-

tence, 'o is G *because* o is F and all F are G', is, if true, contingently true, even though the corresponding assertion of entailment, 'if all F are G and o is F, then o is G', is a necessary truth.

Moreover, any attempt to minimize the difference between an argument theory and a non-argument sentence theory works more to my advantage than to Hempel's. It is a doctrine central to Aristotle's, Mill's, and Hempel's accounts that explanations are arguments. In so far as the distinction between an argument and a sentence is minimized, it is a central doctrine of theirs whose importance is being reduced.

We have, at a sweep, a convincing reason for dismissing any argument theory of explanation, whether deductivist or probabilist. (We still have the choice between certainty, high, and low epistemic probability theories of explanation, the first two being the non-argument analogues of deductivism and probabilism.) In particular, this criticism strikes at the very heart of the Mill-Hempel theory, and the Aristotelian theory of scientific explanation, for all three thinkers held that all full explanations were deductive or inductive arguments. These accounts of explanation not only fail to offer sufficient conditions for full explanation, but more importantly they fail even to provide *necessary* ones. The criticism is not that explanations are not just arguments, but rather arguments plus something more; explanations are, typically, not arguments at all. . . .

If explanations are typically not arguments, what place do laws have in explanation? Can we argue that, since explanations typically are not arguments, therefore explanations typically do not include laws? Although I do believe that many full explanations do not include laws, I do not think that the absence of laws from even some explanations at all follows from the fact that some explanations are not arguments.

The requirement that explanations always include at least one lawlike generalization has been closely bound up with argument theories of explanation. That is to say, if all explanations were deductively valid or inductively good arguments, they would (given the addition of some further uncontroversial assumptions) have to include a lawlike generalization as a premiss.¹⁸ But the inverse is not true; it does not follow from the fact that not all explanations are arguments, that a law is not a part of every full explanation. It only follows that, if laws are a part of full explanations which are not arguments, the idea of their parthood in such cases is not to be cashed out as that of a premiss in an argument. For example, suppose (S) is an explanation of why e happened: (S) 'e occurred because of the fact that c occurred and that whenever a C, an E.' (S) is a sentence, not an argument, and yet it includes the statement of a law.

However, McCarthy's example, in conjunction with my additional remarks about the solution for the difficulty he detects, and Scriven's example below, also convincingly show that laws are not part of every full explanation, in any sense of parthood. The idea that full explanations do

not always include laws (and therefore are not always arguments) is not a novel one. In different ways and from different points of view, Ryle, Scriven, Salmon, and Achinstein (and others too; the list is not intended to be exhaustive) have said this, or similar things, about the role of laws in explanation.¹⁹ For example, in numerous papers, Michael Scriven said things similar to what I would wish to maintain about the role of laws or generalizations in explanation (although I do not need to agree with any of his specific examples). In 'Truisms as the Grounds for Historical Explanations', he defended the view that the following was a perfectly *complete or full* explanation as it stood: the full explanation of why (a) William the Conqueror never invaded Scotland is (b) that 'he had no desire for the lands of the Scottish nobles, and he secured his northern borders by defeating Malcolm, King of Scotland, in battle and exacting homage' (Scriven 1959: 444). The explanation, (b), is a conjunctive statement formed from two singular statements and contains no laws. Explanations which lack laws are 'not incomplete in any sense in which they should be complete, but certainly not including the grounds which we should give if pressed to support them' (p. 446). Notice that Scriven can be taken as making a weaker and a stronger point: (a)'s full explanation, whatever it is, includes or may include no law; (b)—which includes no law—is (a)'s full explanation. I agree with the weaker of Scriven's points; there are some full explanations which do not include laws, and (a)'s full explanation is likely to be such an example. I do not necessarily agree that (b) is (a)'s full explanation. I return to this distinction below.

Scriven's example above is an explanation of a human action. It is sometimes argued in the case of human actions that they are explicable but *anomic* [not governed by laws]. The thought here is rather different. Human actions might be, perhaps must be, *nomio*, law-governed. The first of Scriven's claims is that although or even if human actions are always *nomio*, sometimes the laws or 'truisms' which 'cover' them form no part of their full explanation.²⁰

Scriven makes it clear that he intends the point as a point about explanation generally, not just as a point about the explanation of human action.

... abandoning the need for laws ... such laws are not available even in the physical sciences, and, if they were, would not provide explanations of much interest. ... When scientists were asked to explain the variations in apparent brightness of the orbiting second-stage rocket that launched the first of our artificial satellites, they replied that it was due to its axial rotation and its asymmetry. This explanation ... contains no laws. (Scriven 1959: 445)

I have been arguing that some full explanations do not include laws. But laws are still important, even to those cases of explanation which do

not include them, in other ways. Indeed, the argument view, by insisting that laws are a part of every full explanation, has tended to neglect the other ways in which laws are important to explanation. Let me add some remarks about how laws are still important for the explanation of the world about us, all consistent with my above claim; the remarks will also permit me to sharpen my view somewhat on the role of laws and generalizations in explanation.

First, to repeat what I mentioned above, I have argued that there are some full explanations of which laws form no part, in any sense. But many full explanations do include laws, and this seems to be especially so in the special sciences. Indeed, this is one way in which actual explanations, whether 'ideal' or not, in science and ordinary affairs typically differ. Explanations in science typically include relevant laws, although even when this is so, their inclusion in the explanation will not necessarily be as a major premiss of an argument: 'o is G because o is F and all F are G' is a (contingently true) sentence which includes a law, but is not an argument.

Second, laws are important for the resolution of many types of puzzlement. Clearly, citation of an appropriate regularity can show that the phenomenon about which I may be perplexed or puzzled is, in any case, not atypical or extraordinary or irregular in any way. Given Mill's view of the epistemic circularity of deduction, it was not easy to see why he thought explanations had to be deductive arguments with at least one lawlike premiss.* One line of response I proposed on his behalf was that what a covering law 'explanation' of, for example, the Duke of Wellington's mortality could do, was to show how the good Duke's mortality fits into a pattern of nature; the deductive 'explanation' places his mortality within the context of a wider generalization, and hence within the context of a uniformity of nature. I believe that Mill was thinking along such lines as these, since explanation for him was always the fitting of facts into ever more general patterns of regularity. But the answer that I gave on his behalf invites the following observation: explaining the Duke's mortality is one thing; fitting his mortality into a more general pattern, however worthy that may be, is something else. To learn that something is not irregular is not the same thing as to explain it. Not all resolutions of puzzlement or perplexity are *ipso facto* explanations.

There is a *third* way in which laws can be important. Does the explanans really fully explain the explanandum? Perhaps it is not adequate to explain it fully; something may be missing. How can I justify my claim that the explanans fully does the job it is meant to do? On Scriven's (1959: 446) view, suppose I claim that the full explanation of e is c. If I am

* In this paragraph, Ruben is referring to his discussion of Mill in chapter 4 of *Explaining Explanation*, from which the present reading is taken.

challenged about the adequacy or completeness of my explanation, I can justify my claim to completeness, and thereby rebuff the challenge; by citing a law (or truism), e.g. that all C are E (c being a C; e being an E). This is what Scriven calls the 'role-justifying grounds' that laws provide, in support of a claim that one has given a full explanation. The law or truism can justify my assertion that c is the full and adequate explanation of e, without being part of that explanation. Although Scriven does not say so, there can be no objection to offering the full explanation and the justification for its fullness in a single assertion, but if this is done, we should be clear that what we have is a full explanation and something else, and not just a full explanation.

It is for this reason that I distinguished Scriven's weaker and stronger claims above. I agreed that a full explanation for (a) included no laws, but I did not necessarily agree that (b)—which included no laws—was (a)'s full explanation. The full explanation of o's being G is the fact that o is F, only if it is a law that all F are G, *sans* exception. Suppose the law in question is a more complex law which says: (x) (Fx & Kx & Hx & Jx \supset Gx). A full explanation of why o is G would be the fact that o is F&K&H&J. In this way, my view of full explanation is, in at least one way, close to Hempel's, in spite of my rejection of his, or any, argument theory of explanation. Full explanations, on my view as on his, may well be close to ideal things; if almost no one ever gives one, that tells us a lot about the practical circumstances of explanation-giving, but provides no argument whatsoever against such an account of full explanation.

There may be perfectly good pragmatic reasons why we are entitled to give a partial explanation of o's G-ness; it may be that o's being K&H&J is so obvious, that one never needs to say anything more than that o is F. But the law (or 'truism') provides the criterion for what a complete or full explanation is. I do not want to commit myself about the 'fullness' of Scriven's explanation for William's non-invasion of Scotland, since this raises issues about whether there are any laws which 'cover' human actions and which are also expressible in the vocabulary of human action itself, as Aristotle seemed to believe. This would also involve a discussion of how 'truisms', in Scriven's parlance, differ from laws, and I avoid this issue here.²¹

But, to turn to his second example, I am sure that the explanation of the variations in apparent brightness of the orbiting second-stage rocket that launched America's first artificial satellite, in terms only of its axial rotation and asymmetry, cannot be its full explanation. I agree that its full explanation, whatever it is, need not include a law, but since the explanation Scriven offers fails to contain any particular information about, for instance, the source of light that was present, it could not be a full explanation. Scriven's own remarks about the role-justifying grounds that laws provide helps make this very point. The particular explanation Scriven

offers as full can be seen to be only incomplete, not because it does not include a law, but because the law provides the test for fullness which Scriven's explanation fails.

Fourth, on my view, there is still a connection between singular explanation and generality, but not through the presence of a law. Suppose it is argued that the following is a full explanation:²² (F) object *o* is *G* because *o* is *F*. It seems to me that someone who insists that this cannot be a full explanation because of the absence of a law has to motivate the thought that (F) could not really be a full explanation, by showing what it is that (F) omits, which is not omitted once a law is added to the explanation. (Recall that we are already assuming that argument theories of explanation have been rejected, so he can't fault the absence of the law on the grounds of non-derivability of explanandum from explanans without it.) He must, I think, say this: the real full explanation is only (FL): object *o* is *G* because *o* is *F* and $(x) (Fx \supset Gx)$.

But can we pinpoint what it is that the law is meant to add to (F)? What has (FL) got that (F) lacks? . . . What matters to explanation are properties.²³ When *o*'s being *F* fully explains *o*'s being *G*, it isn't (to put it crudely) that *o*'s being *F* explains *o*'s being *G*; there is nothing special about *o* in any of this. Rather, it is *o*'s being *F* that explains *o*'s being *G*. Explanatory impact is carried by properties and there is generality built into the singular explanation by the properties themselves, without the inclusion of a law. This implicit generality surely implies that other relevantly similar *G*s which are *F* will get the same full explanation that *o* got.

Of course, there is one obvious sense in which an explanation of *o*'s being *G*, in terms of *o*'s being *F*, could be incomplete. The explanation might fail to specify or cite all of the explanatorily relevant properties or characteristics of *o*. But all of the relevant properties of *o* can be cited without inclusion of any law generalization.

Suppose, for the sake of argument, that it is an exceptionless law of nature that $(x) (Fx \supset Gx)$. In this case, the only property of *o*, relevant for explaining why *o* is *G*, is *o*'s *F*-ness. In such a case, it seems that *o*'s being *G* can be fully explained by *o*'s being *F*. What could the inclusion of the law or generalization add to the explanation that *o* is *G* because *o* is *F*?

In ' $(x) (Fx \supset Gx)$ ', the only information that could be relevant to the explanation of *o*'s being *G* is already given by the property linkage between being *F* and being *G* which is already expressed by (F). That part of the information in the generalization which is about (actual or possible) *F*s other than *o* which are also *G*, is simply irrelevant to the explanation of *o*'s being *G*. In short, everything relevant to the explanation of *o*'s being *G* is already contained in (F), since that claim already makes the requisite property connection between being *F* and being *G*. Assuming that the

generalization can connect properties at all (it is unclear that a generalization can do this, even when strengthened by a necessity operator), what (FL) does that is not done by (F) is to extend the connection to cases other than *o*. And this can't have any additional explanatory relevance to *o*'s case. The case of temporally and spatially distant F-objects which are G is surely not relevant to *o*. One might ask about explanation the question Hume asked himself (but believed he could answer) about his constant conjunction theory of causation: 'It may be thought, that what we learn not from one object, we can never learn from a hundred, which are all of the same kind, and are perfectly resembling in every circumstance.'²⁴

My view is even more radical than the suggestion . . . that Mill could have considered a type of real explanation, parallel to his account of the fundamental kind of real, non-deductive inference. Such Millian considerations would certainly dispense with the generalization that all F are G, in the explanation of *o*, which is F, being G. If the manhood of individual persons does not explain their mortality, how could putting all the cases together, as it were, into a generalization, help get explanation off the ground? How could a generalization have some supervenient explanatory power that each instance of the generalization lacks?

Although such a view dispenses with generalizations, it does not dispense with the relevance to *o*'s case of other Fs which are also Gs. This Millian inspired view of explanation would retain, as relevant to the explanation of *o*'s being G, the F-ness and G-ness of other particulars, *a*, *e*, *i*, *u*, etc. Mill thought that we could (really) infer (and, let us suppose, explain) the Duke's mortality, not from a generalization, but from his resemblance to other individual men who were mortal. Yet, it is hard to see how, if the Duke's manhood cannot explain his mortality, introducing the manhood and mortality of people other than the Duke (whether by a generalization or by the enumeration of other particular instances) could explain it. What is the relevance to the good Duke's mortality of the mortality of men spatially and temporally far distant from him?

On my more radical view, neither the generalization that all Fs are Gs, nor the F-ness and G-ness of other particulars, is required to be any part of the full explanation of *o*'s being G. In the case being supposed, the only fact required for the full explanation of *o*'s being G, is *o*'s being F, even though the generalization, and the explanation of other particulars' G-ness by their F-ness, and so on, are implied or presupposed by the full explanation of *o*'s being G by *o*'s being F. The question Hume asked, quoted above, if it has any bite at all, bites not only against a constant conjunction theory of causation (which brings a generalization into prominence), but even against a weaker theory of causation which makes part of the analysis of an instance of a causal relation information about *any other* individual instances of that causal relation.

■ | Generalizations Get Their Revenge

The above remarks attempt to spell out a number of ways in which laws and generalizations are important for explanation, without necessarily being part of them. There is yet another way, closely connected with the third and fourth ways mentioned above. It is sufficiently important to separate it from the others. . . .

Aristotle, it will be recalled, thought that laws provided the criteria for the selection of the descriptions under which the explanans explains the explanandum. Why did the match light? I struck it, and my striking of the match was, let us suppose, the penultimate thing that ever happened to the match. Or, my striking of the match was the event that caused the match to light. Why, then, can I *explain* the fact that the match lit by the fact that the match was struck, and not by the different facts that the penultimate thing that ever happened to the match occurred, or that the cause of its lighting occurred, even though these three singular facts (the fact that the match was struck, the fact that the cause of the match's lighting occurred, the fact that the penultimate thing that ever happened to the match occurred) are all facts about the same causal event, but differently described? In virtue of which of the features of a cause is the cause fully explanatory of the effect?

Aristotle's reply would be that the explanatory features are the ones linked in a law (whether deterministic or stochastic). To be sure, that strikings of matches are followed by lightings of matches is itself no law: nor any part of a law, of nature. We must therefore extend Aristotle's point to include not only features linked in a law, but also features nomically connected in the appropriate way in virtue of underlying laws. . . . In virtue of the underlying laws of physics and chemistry, striking and lighting, but not for example being a penultimate occurrence and a lighting are nomically related. It is not that the laws need be any part of the explanation; rather, the laws provide the criteria for determining under which descriptions one particular explains another (which singular fact explains another). Laws permit selection of the vocabulary appropriate for singular explanation.

The above allows me to make a closely related point about the role of theories in explanation. Scientists often cite theories in explaining phenomenon. For example: the theory of gravity explains why the moon causes the earth's tides; the law of inertia explains why a projectile continues in motion for some time after being thrown; subatomic particle theory explains why specific paths appear in a Wilson cloud chamber. All theories consist (perhaps *inter alia* [among other things]) of generalizations. But (a) it does not follow that theories are explanatory in virtue of their generality, (b) nor does it follow that the way in which they are explanatory is in all cases by being part of the explanation. I have already

argued for (b). But I now wish to argue for (a). Theories help to explain singular facts, in virtue of supplying a vocabulary for identifying or redescribing the particular phenomena or mechanisms at work, which are what explain the explanandum facts.

The examples of 'syllogistic explanation' that I used in my discussion of Mill might have struck the reader as exceedingly artificial: whoever would have thought, the reply might go, that the Duke of Wellington's mortality could be explained by his manhood and the generalization that all men are mortal? And, in admitting that some explanations do include laws (especially in the sciences), I gave this example: 'o is G because o is F and all F are G'. These generalizations are 'flat', in the sense that they are simple generalizations that use the same vocabulary as do the singular explanans and explanandum descriptions. Flat generalizations do not contribute at all to singular explanation.

However, from the fact that flat generalizations are explanatorily useless, it hardly follows that all are. What is needed, so the reply might continue, are generalizations which employ a theoretical vocabulary with greater depth than 'man' and 'mortal'. Perhaps the vocabulary should be in deeper terms that refer to the fragility of hydrocarbon-based life forms. To explain why o is G, in terms of o's being F, if a law is to be included, typically a scientific explanation will cite a law with a vocabulary which is different from and deeper than the vocabulary of which 'F' and 'G' are part. Only as such could the generalizations be explanatory.

And such a reply is correct. But it confirms rather than disconfirms my view. If generalizations or laws were always *per se* explanatory, then flat ones ought to help explain (perhaps not as well as deep ones, but they should explain to some extent none the less). The fact that *only* ones that are deep, relative to the vocabulary of the explanans and explanandum singular sentences (in general, theories), will help explain *at all* is an indication that they are explanatory in virtue of offering a deeper vocabulary in which to identify or redescribe mechanisms, but not just in virtue of being generalizations. And even so, to return for a moment to (b), the generalizations that make up the wider or deeper theory may help to explain by offering that alternative vocabulary, and without being part of the explanation itself.

I argued before that often full explanations do not include laws, but that they sometimes do, especially in the special sciences. When laws are included within an explanation, as they sometimes are, the purpose of the inclusion is to introduce a vocabulary different from the one used in the explicit descriptions of the particular explanans and explanandum events. On the one hand, if the less deep vocabulary used to describe the particular phenomena were wholly expendable, the theoretic vocabulary could be explicitly used to describe them, and any mention of the law would be redundant. If on the other hand no deeper vocabulary were available, there would be no purpose for a law to serve. Laws find their honest employ-

ment in singular explanation in situations between the two extremes: when the less deep vocabulary used to describe singular explanans and explanandum is to be retained at that level, but a deeper vocabulary is available, and needs introduction.

One important role that theories play in science is to unify superficially diverse phenomena.²¹ In virtue of a unifying theory, what seemed like different phenomena can be brought under one set of deep structural laws:

By assuming that gases are composed of tiny molecules subject to the laws of Newtonian mechanics we can explain the Boyle-Charles law for a perfect gas. But this is only a small fraction of our total gain. First, we can explain numerous other laws governing the behavior of gases. . . . Second, and even more important, we can integrate the behavior of gases with the behavior of numerous other kinds of objects. . . . In the absence of the theoretical structure supplied by our molecular model, the behavior of gases simply has no connection at all with these other phenomena. Our picture of the world is much less unified. (Friedman 1981: 7)

On my view, there is a difference between unification and explanation. Unification of a phenomenon with other superficially different phenomena, however worthwhile a goal that may be, is no part of the explanation of that phenomenon. If other men's mortality couldn't explain why the good Duke is mortal when his own manhood doesn't, then the fragility of other hydrocarbon-based life forms couldn't explain the Duke's fragility or mortality when his own hydrocarbon constitution doesn't. It doesn't matter, from the point of view of explanation, whether there are any other phenomena which get explained by the deeper vocabulary; the point is that the vocabulary gives a new and more profound insight into the phenomenon at hand, whether or not the vocabulary unifies it with other phenomena.

■ | Notes

1. Ardon Lyon, 'The Relevance of Wisdom's Work for the Philosophy of Science: A Study of the Concept of Scientific Explanation', in *Wisdom: Twelve Essays*, ed. Renford Bambrough, Blackwell, Oxford, 1974, pp. 218–48.
2. Baruch Brody, 'Towards an Aristotelian Theory of Explanation', *Philosophy of Science*, vol. 39, 1972, pp. 20–31.
3. Peter Achinstein, *The Nature of Explanation*, Oxford University Press, New York, 1983. Discussion of this example on pp. 168 and 170–1.
4. The case of causal pre-emption presents some difficulty for any analysis of deterministic and/or nondeterministic causation which makes a cause necessary

in the circumstances for its effect. . . . David Lewis does not think that a non-deterministic cause is necessary in the circumstances for its effect, but he does think that a deterministic cause is. Lewis deals (on p. 191) with the case of pre-emption in 'Causation', reprinted in Ernest Sosa, *Causation and Conditionals*, Oxford University Press, Oxford, 1975, pp. 180–91; his treatment is discussed by William Goossens, 'Causal Chains and Counterfactuals', *Journal of Philosophy*, 1979, pp. 489–95.

5. Michael Redhead suggests this reply in a paper, 'Explanation' (unpublished): '... we need to attend to all the relevant circumstances. . . . Again the scientific ideal assumes that all the relevant circumstances are being cited' (p. 5). Redhead's reply to my argument is that I neglect the relevant microphysical circumstances linking the bus, but not the arsenic, with the death. On this sort of view, at best only microphysical explanation will meet Hempel's requirements for explanation.

6. In private discussion.

7. Wesley Salmon, 'A Third Dogma of Empiricism', in *Basic Problems in Methodology and Linguistics*, ed. Robert Butts and Jaakko Hintikka, Reidel, Dordrecht, 1977, pp. 149–66. Readers can learn about the nervous husband and the religious explainer on p. 150. Also section 2 of Wesley Salmon, 'Statistical Explanation', in R. Colodny, ed., *The Nature and Function of Scientific Theories*, University of Pittsburgh Press, Pittsburgh, 1970, pp. 173–231, reprinted in W. Salmon, R. Jeffrey, and J. Greeno, *Statistical Explanation and Statistical Relevance*, University of Pittsburgh Press, Pittsburgh, 1971, pp. 29–88.

8. James Woodward, 'Explanatory Asymmetries', *Philosophy of Science*, vol. 51, 1984, pp. 421–42. See also Evan Jobe, 'A Puzzle Concerning D-N Explanation', *Philosophy of Science*, vol. 43, 1976, pp. 542–9, and Clark Glymour, 'Two Flagpoles are More Paradoxical Than One', *Philosophy of Science*, vol. 45, 1978, pp. 118–19. Peter Achinstein, *op. cit.*, p. 236: 'It is possible to explain the presence of a macro-property by appeal to the presence of an identical micro-property; or vice-versa.' Achinstein does not draw the conclusion explicitly that explanation is not asymmetric, but the conclusion follows from what he does say.

9. Aristotle's example of vines which are deciduous because broad-leaved provides a 'symmetry' counterexample to Hempel's account of the explanation of laws.

10. Wesley Salmon, 'A Third Dogma of Empiricism', p. 150.

11. Wesley Salmon, *Scientific Explanation and the Causal Structure of the World*, Princeton University Press, Princeton, 1984, p. 192, p. 96.

12. Baruch Brody, *op. cit.*, pp. 23–4.

13. Timothy McCarthy, 'On an Aristotelian Model of Scientific Explanation', *Philosophy of Science*, vol. 44, 1977, pp. 159–66.

14. Jaegwon Kim 'Discussion: On the Logical Conditions of Deductive Explanation', *Philosophy of Science*, vol. 30, 1963, pp. 286–91. The conjunctive normal form requirement is introduced on p. 288.

15. McCarthy, *op. cit.*, pp. 161–2. Can we strengthen the causal requirements, to rule out a McCarthy-style counterexample? In the arguments so far, the law, even though it might be a causal law, is 'irrelevant' to the explanation (although

it is not irrelevant for the derivation). The law may be a causal law, but it does not join the cause of the explanandum event with the explanandum event. The law premiss and the description of the explanandum's cause don't mesh together. In McCarthy's example, the law (let's assume that it is an irrelevant causal law) was: $(x)(Ax \supset Bx)$, but the description of the explanandum's cause was 'Ce'. As McCarthy says: 'Let " $(x)(Ax \supset Bx)$ " represent any law irrelevant to the occurrence of e' (p. 161). In the second example, the law relates blackness and crows, but the cause of o's turning black has nothing to do with the law; the cause of o's turning black is having been immersed in black paint. No law in the derivation related black paint immersion and turning black. Perhaps a bit of tinkering is all that is needed. Can we impose a further requirement, and thereby escape the counter-example to Hempel's theory as supplemented by the causal requirement?

Suppose we impose the additional requirement that not only must there be a premiss essential to the argument which describes C, the particular cause of the event to be explained, but that there also must be a law premiss essential to the argument such that c(o) and the event to be explained, in this case, o's turning black, are covered by that law. That is to say, the law itself must not be 'irrelevant'; it must bring together the event to be explained and the cause of that event. Thus, the additional 'relevance' needed can be cashed out as 'the law must be a covering law which covers the token cause and effect mentioned in the explanatory argument'. If there is one law which covers both the token cause and the token effect, the law will be a causal law.

We must not require that the explanandum event and the cause be covered by the same law, as the above suggests. This would be too strong, for surely there are occasions on which we can explain an effect by its cause mediately, rather than immediately, via two or more laws. Perhaps we should require that however many laws there are, not only must the premisses contain a description of the cause of the event to be explained, but that both the cause and the explanandum event must be covered by relevant laws, which may relate the cause with the effect only mediately, so that the cause and effect can each be covered by a different law. No doubt, at least one of the laws will be a causal law; but it would be too strong to require that all of the 'interconnecting' laws be causal: I can explain the period of a metal pendulum at t' by citing the fact that heat was applied to the pendulum at t, the causal law that heat causes metal to expand, and the (non-causal) law that relates the length and period of a pendulum.

Even this additional condition will not let us deal with McCarthy's third case, which is as follows. I shall first sketch the third example informally, in order to make it fully intuitive. Suppose o's being F causes o to be G. One would imagine that the explanation of o's being G is o's being F, via the causal law (for let us suppose that it is a causal law) that whatever is F is G. But, with certain other assumptions about the case, we can construct an argument which meets all of the Mill-Hempel conditions, even supplemented in all the required causal ways I have suggested, but which still fails to explain.

As we have already specified, o's being F causes o to be G. What we are to imagine is a case in which the cause of an event to be explained is also the cause of the prediction of that event: If a machine of type T is brought into contact with an object which is F, the object's being F causes the machine to predict that the object is G, as well as causing the object to be G. Moreover, the machines are to

be of type T, which are 'infallible predictors': if it predicts that an object is G, it follows that the object is G. We can now obtain the following argument, noting that (2) states a causal law:

- 1 If a machine is of type T and if it predicts that an object is G, it follows that the object is G.
- 2 If an object is F, and if a machine of type T is in the right relationship with the object, the machine will predict that the object is G.
- 3 Object o is F.
- 4 The machine of type T is in the right relationship with object o.

- 5 Object o is G.

This argument meets all the conditions we have laid down. The premisses include essentially a description of the cause of o's being G, namely o's being F. Further, the premisses include laws which cover and connect the cause and effect, and at least one of which is a causal law. But still, I believe, the argument is not an explanation of why o is G. The object o is G because it is F, and nothing in the derivation reflects this.

16. My suggestion for remedying the difficulty McCarthy points out is taken from, or anyway inspired by, Peter Achinstein, *op. cit.*, pp. 159–62, 188–92.

17. This idea is close to Peter Achinstein's conception of a complete content-giving proposition. I do not believe, though, that any purely grammatical characterization of this idea is possible. See Peter Achinstein, *ibid.*, pp. 28–48, and my review of his book in the *British Journal for the Philosophy of Science*, vol. 37, 1986, pp. 377–84.

18. That there must be a lawlike generalization among the premisses in an explanatory argument does not follow simply from the assumption that explanations are arguments, for there are sound arguments with no such premiss. But the additional assumptions that would be needed in the case of arguments that are explanations are straightforward and uncontroversial to the question at hand.

19. Gilbert Ryle, "If," "So," and "Because", in Max Black, ed., *Philosophical Analysis: A Collection of Essays*, Prentice-Hall, Englewood Cliffs, NJ, 1963, pp. 302–18; Michael Scriven, in a series of contributions, but perhaps especially in 'Truism as the Grounds for Historical Explanations' in *Theories of History*, ed. Patrick Gardiner, Free Press, New York, 1959, pp. 443–75 (see p. 446; page references in the text are to 'Truism . . .'); Wesley Salmon, 'A Third Dogma of Empiricism', pp. 158–9; Peter Achinstein, *op. cit.*, pp. 81–3, and also in 'The Object of Explanation', in *Explanation*, ed. Stephan Körner, Blackwell, Oxford, 1975, pp. 1–45.

20. See Thomas Nickles, 'Davidson on Explanation', *Philosophical Studies*, vol. 31, 1977, pp. 141–5, where the idea that 'strict' covering laws may be 'non-explanatory' is developed.

21. Scriven's distinction is similar to Donald Davidson's, between homonomic and heteronomic generalizations. See Davidson, 'Mental Events', reprinted in his *Essays on Actions and Events*, Oxford University Press, Oxford, 1980, pp. 207–27; see especially pp. 218–20.

22. Note that I say '... that the following is a full explanation'. It is no part of my view that there can be at most only one full explanation for a singular fact. To take just one possibility: suppose one wants to explain why *o* is *G*. Suppose it is a law that all *D* are *F*, and a law that all *F* are *G*. The fact that *o* is *G* can be fully explained both by the fact that *o* is *F*, and the fact that *o* is *D*.

23. I do not deny that there can be cases of explanation in which explanatory relevance is borne by names. . . . But I do not deal with these cases here.

24. David Hume, *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, Oxford University Press, Oxford, 1965, p. 88. I think that much of the motivation for the inclusion of a generalization in every full explanation stems from the Humeian analysis of causation.

25. See Michael Friedman, 'Theoretical Explanation', in *Reduction, Time, and Reality*, ed. Richard Healey, Cambridge University Press, Cambridge, 1981, pp. 1-16. See also his 'Explanation and Scientific Understanding', *Journal of Philosophy*, vol. LXXI, 1974, pp. 5-19, and the reply by Philip Kitcher, 'Explanation, Conjunction, and Unification', *Journal of Philosophy*, vol. LXXIII, 1976, pp. 207-12.

A Deductive-Nomological Model of Probabilistic Explanation

What if some things happen by chance—can they nonetheless be explained? How?

Some things *do* happen by chance, according to the dominant interpretation of our present physical theory, the probabilistic interpretation of quantum mechanics. Nonetheless, they can be explained by that theory, in virtually the same way as deterministic phenomena—deductive-nomologically. At least, that is what I hope to show in this essay.

Our universe may not be deterministic, but all is not chaos. It is governed by laws of two kinds: probabilistic (such as the laws concerning barrier penetration and certain other quantum phenomena) and non-probabilistic (such as the laws of conservation of mass-energy, charge, momentum, etc.).¹ Were the probabilism of laws of the first sort remediable by suitable elaboration of laws of the second sort, the universe would be deterministic after all, and the problem of explaining chance phenomena would no longer be with us. However, indications are that physical indeterminism is irremediable, and that the universe exhibits not only chances, but lawful chances. I will argue that we come to understand chance phenomena, even when the chance involved is extremely remote, by subsuming them under these irremediably probabilistic laws.

1 | Introductory Remarks on Explanation

Do I offer a deductive-nomological (D-N) model of probabilistic explanation because I believe that nomic subsumption always explains?—No. There are familiar-enough kinds of non-explanatory D-N arguments, for example, those that deduce the explanandum from nomically-related symptoms or after-the-fact conditions alone, citing no causes.

Yet it will not do simply to add to the D-N model a requirement that the explanans contain causes whenever the explanandum is a particular fact. First, some particular facts may be explained non-causally, for example, by subsumption under structural laws such as the Pauli exclusion principle.⁸ Second, even where causal explanation is called for, the existence of general, causal laws that cover the explanandum has not always been sufficient for explanation: the search for explanation has also taken the form of a search for mechanisms that underlie these laws. 'Mechanisms,' however, is not meant to suggest a parochial attitude toward the nomic connections—deterministic or otherwise—that tie the world together and make explanation possible.

An example may help clarify the notion of mechanism appealed to here. The following D-N argument suffices to forecast *that* nasty weather lies ahead, but not to explain *why* this is so

S The glass is falling.

Whenever the glass falls the weather turns bad.

The weather will turn bad ([5], p. 106)

Now nothing works like a barometer for predicting the weather, but nothing like a barometer works for changing it. So it is often maintained that (S) lacks explanatory efficacy because barometers lack the appropriate causal efficacy. The following inference, then, remedies the lack of the first because "it proves that the fact is a fact by citing causes and not mere symptoms" ([5], p. 107):

C The glass is falling.

Whenever the glass is falling the atmospheric pressure is falling.

Whenever the atmospheric pressure is falling the weather turns bad.

The weather will turn bad. ([5], p. 106)

⁸ The Pauli exclusion principle was discovered in 1925 by Wolfgang Pauli (1900–58). It says that no two electrons can have the same set of quantum numbers or, in other words, that no two electrons can be in the same quantum state. Originally formulated for electrons, the exclusion principle is now known to apply to all fermions (particles having half-integer spin). Among other things, it explains why the electrons surrounding the nuclei of different elements are arranged in shells rather than all occupying the state of lowest energy closest to the nucleus. As Railton points out, this kind of explanation is not causal: the principle is a purely formal or structural law. Other structural laws include the conservation principles of energy and momentum, Einstein's principle that the velocity of light is constant in all inertial frames, and the second law of thermodynamics.

Yet as explanations go, (C) is also lacking: we remain in the dark as to *why* the weather will turn bad. No connection between cause and effect, no mechanism by which falling atmospheric pressure produces a change for the worse in the weather, has been revealed. I do not doubt that some account of this mechanism exists; my point is that its existence is what makes (C) superior to (S) for explanatory purposes.

(C), if moderated by boundary conditions and put less qualitatively, would supply us the capability to predict *and* control the weather (when-ever, as in a laboratory simulator, we can manipulate the atmospheric pressure). While prediction and control may exhaust our practical problems in the natural world, the unsatisfactoriness of (C) shows that explanation is an activity not wholly practical in purpose. The goal of understanding the world is a theoretical goal, and if the world is a machine—a vast arrangement of nomic connections—then our theory ought to give us some insight into the structure and workings of the mechanism, above and beyond the capability of predicting and controlling its outcomes. Until supplemented with an account of the nomic links connecting changes in atmospheric pressure to changes in the weather, (C) will explain but poorly. Knowing enough to subsume an event under the right kind of laws is not, therefore, tantamount to knowing the *how* or *why* of it. As the explanatory inadequacies of successful practical disciplines remind us: explanations must be more than potentially-predictive inferences or law-invoking recipes.

Is the deductive-nomological model of explanation therefore unacceptable?—No, just incomplete. Calling for an account of the mechanism leaves open the nature of that account, and as far as I can see, the model explanations offered in scientific texts are D-N when complete, D-N sketches when not. What is being urged is that D-N explanations making use of true, general, causal laws may legitimately be regarded as unsatisfactory unless we can back them up with an account of the mechanism(s) at work. “An account of the mechanism(s)” is a vague notion, and one obviously admitting of degrees of thoroughness, but I will not have much to say here by way of demystification. If one sees what is lacking in (C)—a characterization, whether sketchy or blow-by-blow, of how it is that declining atmospheric pressure effects the changes we describe as “a worsening of the weather,” that is, a more or less complete filling-in of the links in the causal chains—one has the rough idea.

The D-N probabilistic explanations to be given below do not explain by giving a deductive argument terminating in the explanandum, for it will be a matter of chance, resisting all but *ex post facto* demonstration. Rather, these explanations subsume a fact in the sense of giving a D-N account of the chance mechanism responsible for it, and showing that our theory implies the existence of some physical possibility, however small, that this mechanism will produce the explanandum in the circumstances given. I hope the remarks just made about the importance of re-

vealing mechanisms have eased the way for an account of probabilistic explanation that focuses on the indeterministic mechanisms at work, rather than the "nomic expectability" of the explanandum.

2 | Hempel's Inductive-Statistical Model

For Hempel, a statistical explanation (what is called elsewhere in this paper 'a probabilistic explanation') is one that "makes essential use of at least one law or theoretical principle of statistical form" ([3], p. 380). Since Hempel distinguishes between statistical laws and mere statistical generalizations, and asserts that the former apply only where "peculiar, namely probabilistic, modes of connection" exist among the phenomena ([3], p. 377), his characterization permits statistical explanation only of genuinely indeterministic processes.² Were some process to have the appearance of indeterminism owing to arcane workings or uncontrolled initial conditions, then no "peculiar . . . probabilistic" modes of connection would figure essentially in explaining this "pseudo-random" process's outcomes. Not only would statistical explanation be unnecessary for such a process, it would be impossible: no probabilistic *laws* would govern it.

For example, it has been observed that 99% of all cases of infectious mononucleosis involve lymph-gland swelling. The exceptions might be due to a process that randomly misfires 1% of the time. Or, they might arise from the operation of an unknown deterministic mechanism that works to inhibit swelling whenever a patient begins in a particular initial condition, which as a mere matter of fact is typical of 1% of the population. If initial conditions could be partitioned into two mutually exclusive and jointly exhaustive classes *S* and $\neg S$, such that all *S*s by law eventually develop swelling, and all $\neg S$ s do not, the generalization "99% of all cases of infectious mononucleosis develop lymph-gland swelling" would have been shown to be no law, but merely a descriptive report of observed relative frequencies. No law, it cannot support a statistical explanation. But discovering it not to be a law is just discovering that statistical explanation is uncalled for, since each case of mononucleosis will have been of type *S* or type $\neg S$ from the outset.

On the other hand, suppose that no such partition of initial conditions exists. Then the presence or absence of swelling is presumably due to a "peculiar . . . probabilistic" connection between disease and symptom, that is, a real causal indeterminism with probability .99 in each case to produce swelling. The generalization in question would thus be nomological, creating both the possibility and the necessity of statistical explanation.

Given such genuine statistical laws, how does Hempel claim statistical explanation should proceed? He begins his account by distinguishing two

sorts of statistical explanation. The first, *deductive-statistical* (D-S) explanation, involves "the deductive subsumption of a narrow statistical uniformity under more comprehensive ones" ([3], p. 380). The second, he argues, is of a qualitatively different sort:

Ultimately . . . statistical laws are meant to be applied to particular occurrences and to establish explanatory and predictive connections among them ([3], p. 381)

To make such laws relevant to "particular occurrences," Hempel believes we must go beyond the reach of deduction, and so he proposes an inductive model of statistical explanation.

Inductive-statistical (I-S) explanation proceeds by adducing statistical laws and associated initial conditions relative to which the explanandum is highly probable. High relative probability is required because, on Hempel's view, statistical laws become explanatorily relevant to an individual chance event only by giving us a basis upon which to inductively infer its occurrence with "practical certainty." Yet although an I-S explanation shows the explanandum to have been "nomically expectable" relative to the explanans, it does not permit detachment of a conclusion; it is less an inference than the expression of an inferential relationship: the explanandum receives a high degree of epistemic support from the explanans. If, for example, we learn that Jones has contracted infectious mononucleosis, we may infer with practical certainty that he will develop lymph-gland swelling. The same inference serves as an I-S explanation of the swelling, should it occur. Should it not occur, we would have no explanation for *this*, on Hempel's model.

However, further investigation of Jones' medical history might reveal that he suffered mononucleosis once before, and failed to develop any swelling. Let us suppose that such individuals have a much higher than normal probability of *not* showing swelling in any later bouts with mononucleosis, say .9 rather than .01. This new law and new information about Jones together permit an inference with practical certainty to the conclusion that he will *not* develop swelling, and thus support a corresponding I-S explanation. Relative to these new facts, however, no I-S explanation would be available should Jones, improbably, develop swelling. What are we to say now about the previous I-S explanation, which had just the opposite result? Hempel would reject it as no longer *maximally specific* relative to what we believe about Jones' case. The requirement of maximal specificity is a complicated affair,³ but the basic idea is that we refer each case to the narrowest class of cases to which our present beliefs assign it in which the explanandum has a characteristically different probability. In Jones' case, the narrower class is clearly the class of those contracting mononucleosis for a second time who failed to develop lymph-gland swelling the first time.

If more information about Jones or new discoveries about mononucleosis turn up, we may be forced to move on to still another explanation. I-S explanations must be relativized to our current "epistemic situation," and are subject to change along with it. Hempel notes that this sets off I-S explanations from D-N and D-S explanations in a fundamental way:

... the concept of statistical explanation for particular events is essentially relative to a given knowledge situation as represented by a class K of accepted statements. ... [W]e can significantly speak of true D-N and D-S explanations: they are those potential D-N and D-S explanations whose premises (and hence also conclusions) are true—no matter whether this happens to be known or believed, and thus no matter whether the premises are included in K. But this idea has no significant analogue for I-S explanation. ... ([3], pp 402–3)

On Hempel's view, neither of the two contradictory explanations concerning Jones contains false premises, and the explananda in each case do indeed receive the degree of support indicated. It is just that we no longer regard the evidential relationship expressed by the first as explanatorily relevant. Were Jones to develop swelling after all, it would now have to be regarded as inexplicable.

What I take to be the two most bothersome features of I-S arguments as models for statistical explanation—the requirement of high probability and the explicit relativization to our present epistemic situation (bringing with it an exclusion of questions about the truth of I-S explanations)—derive from the inductive character of such inferences, not from the nature of statistical explanation itself. If a non-inductive model for the statistical explanation of particular facts is given, there need be no temptation to require high probability or exclude truth.

3 | Jeffrey's Criticism of I-S Explanation

Richard C. Jeffrey has criticized Hempel's account on the grounds that statistical explanation is not a form of inference at all, except when the probability of the explanandum is "so high as to allow us to reason, in any decision problem, as if its probability were 1" ([5], p. 105). For such exceptional, "beautiful" cases, Jeffrey accepts I-S inferences as explanatory because they provide virtual "proof that the phenomenon *does* take place" ([5], p. 106).

For unbeautiful cases, there is no way of proving (in advance) that the explanandum phenomenon will occur. According to Jeffrey, the explanation *why* such unbeauties come to be is a curt "By chance." He has more to say on *how* they come about:

. . . in the statistical case I find it strained to speak of knowledge *why* the outcome is such-and-such. I would rather speak of *understanding the process*, for the explanation is the same no matter what the outcome: it consists of a statement that the process is a stochastic one, following such-and-such a law,⁴ ([5], p. 24)

Jeffrey is surely right, as against Hempel, that probable and improbable outcomes of indeterministic processes are equally explicable, and explicable in the same way. After all, why should it be explicable that a genuinely random wheel of fortune with 99 red stops and 1 black stop came to a halt on red, but inexplicable that it halted on black? Worse, on Hempel's view, halting at any *particular* stop would be inexplicable, even though the wheel must halt at some particular stop in order to yield the explicable outcome *red*.

But I fail to see how Jeffrey can defend his exemption of beautiful cases against a similar line of argument. If the burden in statistical explanation really lies with "*understanding the process . . . no matter what the outcome*," then why should it matter whether the outcome is so highly probable "as to allow us to reason, in *any* decision problem, as if its probability were 1?" The neglect Jeffrey shows here toward minute chances is appropriate for the practical task of decision-making (and perhaps explained by his generally subjectivist approach to probability), but we must not overlook them in the not-entirely-practical task of explaining. Virtually impossible events may occur, and they deserve and can receive the same explanation as the merely improbable or the virtually certain.

4 | A D-N Model of Probabilistic Explanation

I will present my account of probabilistic explanation by developing an example of just such "practically negligible"—but physically real and lawful—chance: alpha-decay in long-lived radioactive elements. The mean-life of the more stable radionuclides is so long as to make the probability for any particular nucleus of such an element to decay during our lifetimes effectively zero. But our nuclear theory shows that it is *not* zero, and explains how such rarities can occur.

On the account offered here, probabilistic explanations will be either true or false independent of our epistemic situation. Moreover, to explain, they must be true. Here I am following Hempel's usage in calling an explanatory argument *true* just in case it is valid and its premises are true ([3], p. 338). Such an explanation will *not* be true if the probabilistic laws it invokes are not true; in particular, it will not be true unless the process responsible for the explanandum is genuinely indeterministic. If alpha-decay is to serve as our paradigm for probabilistic explanation, we must

be correct in assuming that the probabilistic wave-mechanical account of particle transmission through the nuclear potential barrier tells us all there is to know about the cause of alpha-decay. At least, it must be true that there are no hidden variables characterizing unknown initial conditions that suffice to account for alpha-decay deterministically. However, I take it to be uncontroversial that alpha-decay is an indeterministic process, if any is.

Let us suppose that we are given an individual instance of alpha-decay to explain: a nucleus of radionuclide uranium²³⁸, call it '*u*', has emitted an alpha-particle during the time interval lasting from t_0 to $t_0 + \theta$, where θ is very small and expressed in standard units. Since the mean-life of U²³⁸ is 6.5×10^9 years, the probability of observing a decay by *u* during this interval is exceedingly small, but unquestionably exists (witness the decay). This probability can be given precisely by using the radioactive decay constant λ_{238} characteristic of all atoms of U²³⁸. Significantly, we need not know when in the course of the history of *u* time t_0 occurs: the probability of decay is unaffected by the age of the atom. Therefore, as long as decay has not yet occurred, individual "trials"—consisting of observing a single isolated radioactive nucleus for successive intervals of the same length—are statistically independent. Using these two facts we can determine the probability of decay for individual nuclei during any time interval chosen: it will be 1 minus the probability that any such nucleus *survives* the interval intact; for *u*, $(1 - \exp(-\lambda_{238} \cdot \theta))$.*

To obtain experimental confirmation of this value, we infer *from* the probability to decay of individual nuclei to statistical features of sample populations of nuclei, for example, half-life and mean-life. These predicted statistical features are then checked against actual observed relative frequencies in large populations over long intervals. *Physical* probabilities of the sort being considered here are therefore to be contrasted with *statistical* probabilities; the former express the strength of a certain physical possibility for a given system, while the latter reduce to claims about the (limiting) relative frequencies of traits in sample populations. Much well-founded doubt has been expressed about the applicability of statistical probabilities to single cases, but physical probabilities are *located* in the features of the single case. Therefore, we can understand our nuclear theory as implying strictly universal (physical) probability-attributing laws of the form:

- (1) All nuclei of radioelement *E* have probability $(1 - \exp(-\lambda_E \cdot t))$ to emit an alpha-particle during any time interval of length *t*, unless subjected to environmental radiation.

* In this exponential decay formula, *exp* stands for *e*, the base of natural logarithms. Thus, the formula should be read as: 1 minus *e* raised to the power of minus lambda times theta.

Because schema (1) is universal in form, its instances are candidates for law premises in deductive-nomological inferences concerning individual nuclei. Thus, for u :

- 2 a All nuclei of U^{238} have probability $(1 - \exp(-\lambda_{238} \cdot \theta))$ to emit an alpha-particle during any interval of length θ , unless subjected to environmental radiation.
- b u was a nucleus of U^{238} at time t_0 , and was subjected to no environmental radiation before or during the interval $t_0 - (t_0 + \theta)$

- c u had probability $(1 - \exp(-\lambda_{238} \cdot \theta))$ to emit an alpha-particle during the interval $t_0 - (t_0 + \theta)$.

(2), it appears, gives a D-N explanation only of the fact that u had such-and-such a probability to decay during the interval in question, but we should look a bit closer. I submit that (2), when supplemented as follows, is the probabilistic explanation of u 's decay:

- 3 A derivation of (2a) from our theoretical account of the mechanism at work in alpha-decay.
The D-N inference (2).
A parenthetic addendum to the effect that u did alpha-decay during the interval $t_0 - (t_0 + \theta)$.

Am I merely making a virtue of necessity, and saying that since (3) contains all we can say about u 's decay, (3) must explain it? In fact, there is a great deal more we could say about u 's decay. Deliberately left out of (3) are innumerable details about the experimental apparatus (temperature, pressure, location, etc.), about the beliefs and expectations of those monitoring the experiment, and about the epistemic position of the scientific community at the time. These facts are omitted as *explanatorily irrelevant* to u 's decay because they are *causally irrelevant* to the physical possibility for decay that obtained during the interval in question, and to whether or not that possibility was realized.⁵ A full account of these notions of explanatory and causal relevance is not possible here, so instead I will go on to argue that what (3) comprises is explanatorily relevant, and explanatory.

I must begin this task with a defense of the nomological status of (2a), and of the legitimacy of treating it as a covering law for u 's decay. The following criterion of nomologicality will be used: a law is a universal truth derivable from our theory without appeal to particular facts. This criterion of course lacks generality (what counts as theory if not the laws themselves?), fails to segregate natural from logical laws, picks out only so-called "universal" (as opposed to "local") laws, and is entirely too vague

how to distinguish "particular facts" from the rest?) But I trust it will do for now. The motive for excluding "particular facts" is that some true, universal statements derivable from our theory *plus* particular facts would not normally be regarded as universal laws, but would at best be "local laws," for example, "All *Homo neanderthalensis* live during the late Pleistocene age."

The generalization in question here, (2a), is derived by solving the Schrödinger wave equation for an alpha-particle of energy ≈ 4.2 MeV for the potential regions in and around the nucleus of an element with atomic number 92 and atomic weight 238, none of which are "particular facts," plus some simplifying assumptions about the structure of the nucleus and the distinctness of the alpha-particle within it prior to decay. While it is forbidden by classical physics for a low-energy particle like the ≈ 4.2 MeV alpha-particle associated with U^{238} to pass through the 24.2 MeV potential barrier surrounding so massive a nucleus, the quantum theory predicts that the probability amplitude for finding such an alpha-particle outside the potential barrier is non-zero. Thus a transmission coefficient for U^{238} alpha-particles is determined, which, given certain simplifying assumptions about the goings-on inside the nucleus, yields the probability that such a particle will tunnel out of the potential well "per unit time for one nucleus," namely, λ_{238} ([1], p. 175). (2a) thus neither reports a summary of past observations nor expresses a mere statistical uniformity that scattered initial conditions would lead us to anticipate. Instead, it is a law of irreducibly probabilistic form, assigning definite, physically determined probabilities to individual systems.

It follows that the derivation of conclusions from (2a) by universal instantiation and *modus ponens* is unexceptionable.* Were (2a) but a statistical generalization, properly understood as meaning " $(1 - \exp(-\lambda_{238} \cdot \theta))N$ of U^{238} nuclei in samples of sufficiently large size N , on average, decay during the interval $t_0 - (t_0 + \theta)$," it could not undergo universal instantiation, and would not permit detachment of a conclusion about the probability obtaining in a single case.

Further, if the wave equation does indeed tell us all there is to know about the mechanism involved in nuclear barrier penetration, it follows that nothing more can be said to explain why the observed decay of u took place, once we have shown how (2a) is derived from our account of this mechanism, and established that (2) is valid and that (3)'s parenthetical addendum is true.

Still, does (3) explain why the decay took place? It does not explain why the decay *had* to take place, nor does it explain why the decay *could be expected* to take place. And a good thing, too: there is no *had to* or *could be expected to* about the decay to explain—it is not only a chance

* See the discussion of this derivation on page 795.

event, but a very improbable one. (3) does explain why the decay *improbably* took place, which is how it did. (3) accomplishes this by demonstrating that there existed at the time a small but definite physical possibility of decay, and noting that, by chance, this possibility was realized. The derivation of (2a) that begins (3) shows, by assimilating alpha-decay to the chance process of potential barrier tunneling, how this possibility comes to exist. If alpha-decays are chance phenomena of the sort described, then once our theory has achieved all that (3) involves, it has explained them to the hilt, however unsettling this may be to *a priori* intuitions. To insist upon stricter subsumption of the explanandum is not merely to demand what (alas) cannot be, but what decidedly should not be: sufficient reason that one probability rather than another be realized, that is, chances without chance.

Because of the peculiar nature of chance phenomena, it is explanatorily relevant whether the probability in question was realized, even though there is no before-the-fact explanatory *argument*, deductive or inductive, to this conclusion. Indeed, it is the absence of such an argument that makes a place in probabilistic explanation for a parenthetic addendum concerning whether the possibility became actual in the circumstances given. These addenda may offend those who believe that explanations must always be arguments, but at the most general level explanations are *accounts*, not arguments. It so happens that for deterministic phenomena inferences of a particular kind—D-N arguments meeting the desiderata suggested in section 1—are explanatory accounts, and this for good reasons. However, indeterministic phenomena are a different matter, and explanatory accounts of them must be different as well. If the present model is accepted, then almost all of the explanatory burden in probabilistic explanation can be placed on deductive arguments—those characterizing the indeterministic mechanism and those attributing a certain probability to the explanandum. But these arguments leave out a crucial part of the story: did the chance fact obtain?

The parenthetic addendum fills this gap in the account, and communicates information that is relevant to the causal origin of the explanandum by telling us that it came about as the realization of a particular physical possibility. Further, it permits us to chain probabilistic explanations together to make more comprehensive explanations, in which each link is able to bear the full explanatory burden for the fact it covers, and is capable of leading us on to the next fact in the causal sequence being explained. From (2) alone we cannot move directly to an account of what the alpha-particle did to a nearby photographic plate, but only to a probability (and a miserably low one) that this account will be true. The parenthetic addendum to (3) furnishes a non-probabilistic premise from which to begin an account of the condition of the photographic plate: the occurrence of an alpha-decay in the vicinity. Dropping off the addendum leaves an explanation, but it is a D-N explanation of the occurrence of a

particular probability, not a probabilistic explanation of the occurrence of a particular decay.

The scheme for probabilistic explanation of particular chance facts by nomic subsumption that is being offered here, the *deductive-nomological-probabilistic* (D-N-P) model, is this. First we display (or truthfully claim an ability to display) a derivation from our theory of a law of essentially probabilistic form, complete with an account of how the law applies to the deterministic process in question. The derived law is of the form.*

$$4a \quad (t)(x)[F_{x,t} \rightarrow \text{Prob}(G)_{x,t} = p]$$

"At any time, anything that is *F* has probability *p* to be *G*."

Next, we adduce the relevant fact(s) about the case at hand, *e*:

$$4b \quad F_{e,t_0}$$

"*e* is *F* at time *t*₀,"

and draw the obvious conclusion:

$$4c \quad \text{Prob}(G)_{e,t_0} = p$$

"*e* has probability *p* to be *G* at time *t*₀."

To which we add parenthetically, and according to how things turn out

$$4d \quad (G_{e,t_0} / \neg G_{e,t_0})$$

"(*e* did/did not become *G* at *t*₀)."

Whether a D-N-P explanation is true will depend solely upon the truth-values of its premises and addendum, and the validity of its logic. I leave open what becomes of a D-N-P explanation that contains true laws, initial facts, and addendum, but botches the theoretical account of the laws invoked. Let us simply say that the more botched, the less satisfactory the explanation.

The law premise (4a) will be true if all things at all times satisfy the conditional ' $F_{x,t} \rightarrow \text{Prob}(G)_{x,t} = p$ ', using whatever reading of ' \rightarrow ' we decide upon for the analysis of natural laws in general. It will be false if there exists a partition of the *F*s into those with *physical* probability *r* to be *G* and those with *physical* probability *s* to be *G*, where $s \neq r \neq p$. Such a partition might exist according to some *other* interpretation of probability, but this would not affect the truth of (4a). For example, suppose that a coin toss meeting certain specifications is an indeterministic event with probability 1/2 of yielding heads. We now perform the experiment of repeating such a toss a great many times. Curiously, all and only

* We have changed Railton's notation for the universal quantifier.

even numbered tosses yield heads. This result supplies certain frequentists with grounds for saying that $\text{Prob}(\text{heads, even-numbered toss}) = 1$, while $\text{Prob}(\text{heads, odd-numbered toss}) = 0$.⁶ But because all tosses met the specification laid down, the probability of heads was the same, $1/2$, on each toss, despite the curious behavior. Such behavior may make us suspicious of our original claims about the indeterminacy of the process or about the physical probability it has of producing heads, but is no proof against them. Indeed, the original probability attribution requires us to assign a definite physical probability to just such an untoward sequence of outcomes, the occurrence of which therefore hardly contradicts this attribution.

The particular fact premise (4b) will be true iff [if and only if] e is an F during the time in question, and not either an F^* (with probability $r \neq p$ to be G) or an F^{**} (with probability $q = p$ to be G , but unlike an F in other respects). Using the (let us say) true law that all F^{**} s have probability $q = p$ to be G , and the falsehood that e is an F^{**} , we could derive a true conclusion, indistinguishable from (4c). Hence the requirement that the *premises* be true if the argument is to explain, and if we reason logically from true premises, the conclusion will take care of itself.

5 | Epistemic Relativity and Maximal Specificity Disowned

Have I kept my promise to give an account of probabilistic explanation free from relativization to our present epistemic situation?

Let us return to explanation (3), and admit that it is not the whole story: 23% of the alpha-particles emitted by U^{238} have kinetic energy 4.13 MeV, while the remaining 77% have 4.18 MeV. Therefore there are two different decay constants, $\lambda_{238}^{4.13}$ and $\lambda_{238}^{4.18}$; both are distinct from λ_{238} used in (3). Hence we must be quite careful in stating what exactly (3) explains. It does *not* explain the particular *event* observed, for this was either a 4.13 or a 4.18 MeV decay, neither of which has probability λ_{238} in unit time. Instead, (3) explains the particular *fact about* the event observed that we set out to explain, namely, that an alpha-decay with unspecified energy (or direction, or angular momentum, etc.) took place at nucleus u during the time interval in question. This fact *does* have probability λ_{238} of obtaining in unit time, representing the sum of the two energy-correlated probabilities with which such a decay might occur.

If we should learn that the decay of u was of a 4.18 MeV alpha-particle, an explanation of *this* fact would have to be referred to the more specific class of decays with probability $\lambda_{238}^{4.18}$ in unit time. Is the maximal specificity requirement thereby resurrected? There is no need for it. (3) is not an unspecific explanation of this more specific fact, but a fallacious

one. It would be logically corrupt to conclude from law (2a) that an individual U^{238} nucleus has probability $(1 - \exp(-\lambda_{238} \cdot \theta))$ to decay with energy 4.18 MeV during any interval of length θ , since (2a) says nothing whatsoever about decay energies. The only relevant conclusion to draw from (2a) is (2c), which remains true in the face of our more detailed knowledge about the event in question. Nor is law (2a) falsified by the discovery of a 23:77 proportional distribution of decay energies, and the associated difference in decay rates. For according to our nuclear theory, there is no difference in initial condition between a nucleus about to emit a 4.13 MeV alpha-particle and one about to emit a 4.18 MeV alpha-particle. It remains true that *all* U^{238} nuclei have probability λ_{238} to decay in unit time, but it is further true that all have probability $\lambda_{238}^{4.13}$ to decay one way, and probability $\lambda_{238}^{4.18}$ to decay another.

It must next be determined whether the existence of a difference in probability *due to* a difference in initial condition can be handled by the D-N-P model without appeal to a maximal specificity requirement. To permit consideration of possible epistemological complications, it will be assumed that neither the difference in probability nor the partition of initial conditions is known at the start.

Imagine that, although we do not know it, in virtue of certain permanent structural features 23% of all naturally-occurring U^{238} nuclei fall into a class P , and the remaining 77% into a class $-P$, such that only those in P have any probability of emitting a 4.13 MeV alpha-particle, and only those in class $-P$ have any probability of emitting a 4.18 MeV alpha-particle. Suppose further that these two laws have been derived:

- 5 a All U^{238} nuclei of type P have probability $(1 - \exp(-\lambda_{238}^{4.13} \cdot t))$ to emit a 4.13 MeV alpha-particle during any time interval of length t , unless subjected to environmental radiation.
- b All U^{238} nuclei of type $-P$ have probability $(1 - \exp(-\lambda_{238}^{4.18} \cdot t))$ to emit a 4.18 MeV alpha-particle during any time interval of length t , unless subjected to environmental radiation.

Note that, by our assumptions, the specification of the kinetic energy of the particle (possibly) emitted may be dropped from (5a) and (5b) without altering the truth of either.

Until the structural differences between types P and $-P$ are discovered and understood, (3) will stand as the accepted explanation of u 's decay. However, once (5a) and (5b) have become known, it will be clear from the fact that u 's alpha-emission had kinetic energy 4.18 MeV that u must have been of type $-P$ prior to decay. Thus a more specific account of u 's decay will be available to scientists, who, already familiar with the theoretical derivation of law (5b), offer the following truncated D-N-P version of this account:

- 6 a All nuclei of U^{238} of type $-P$ have probability $(1 - \exp(-\lambda_{238}^{+12} \cdot \theta))$ to emit an alpha-particle during any time interval of length θ , unless subjected to environmental radiation.
- b u was a nucleus of U^{238} of type $-P$ at t_0 , and was subjected to no environmental radiation before or during the interval $t_0 - (t_0 + \theta)$.

- c u had probability $(1 - \exp(-\lambda_{238}^{+12} \cdot \theta))$ to emit an alpha-particle during the interval $t_0 - (t_0 + \theta)$.
- d (And it did.)

On the Hempelian model (modified so as to permit I-S explanations of improbable phenomena), there is no problem in accounting for the previous acceptability of the I-S counterpart of (3), or for its present unacceptability. (3) had been maximally specific relative to our previous beliefs about alpha-decay in U^{238} , but no longer is, and so is superseded by the more specific (relatively speaking) I-S counterpart of (6).

On the D-N-P model, too, there is no problem in accounting for the acceptability of (3) prior to the discovery of class $-P$ and law (5b): (3)'s premises (and, of course, addendum) were taken to be true. The question is whether, in light of current beliefs, (3) can be ruled out—and (6) ruled in—without invocation of Hempelian constraints. Resolution of the problem (3) and (6) pose through epistemic relativization and maximal specificity requirements seems to me unacceptable. If we were to attribute to nucleus u two unequal probabilities to alpha-decay in a specified way during a single time interval, adding, "Let's pick the most specifically defined value for explanatory purposes," we'd be showing an unseemly tolerance for contradiction in our nuclear theory—and why stop there? Better face up to the confrontation over truth between (3) and (6), and replace complex and unappealingly relativistic maximal specificity requirements with the simple requirement of truth. The D-N-P model does this. The current unacceptability of (3) is located not in premises insufficiently specific, but in premises insufficiently true, that is, false. Contrary to (3)'s purported covering law (2a), not all nuclei of U^{238} have probability λ_{238} to decay in unit time if unperturbed by radiation—in fact, none do. In spite of giving accurate expectation values for decay rates in large samples of U^{238} , (2a) is false, and so explanation (3) is ruled out as unsound. Explanation (6), on the other hand, meets the simple requirement of truth, and rules itself in.⁷

Problems about incomplete, misleading, or false beliefs do not bear on whether D-N-P explanations have unrelativized truth-values, but concern rather difficulties in *establishing* the truth-values they unrelativistically have. Relativization to our current epistemic situation comes into play only when we begin to discuss whether a given D-N-P explanation *seems* true. Whether it is true is another matter.

6 | Objections to the D-N-P Model

I cannot pretend to have said enough about deductive-nomological-probabilistic explanation to have characterized this model adequately. Such reservations as were expressed in section 1 about taking nomic subsumption under a causal law as sufficient for explanation are still in force, and little has been done—except by way of example—to show how the account offered here might accommodate them.

That the probabilistic laws invoked in D-N-P explanations are even (in some relevant sense) *causal* cannot be defended until a plausible account of physical probability has been worked out, a task well beyond the scope of this paper. Under a *propensity interpretation*, probability has the characteristics sought: a probability is the expression of the strength of a physical tendency in an individual chance system to produce a particular outcome; it is therefore straightforwardly applicable to single cases, and it is (in a relevant sense) causally responsible for that outcome whenever it is realized. However, propensities are notoriously unclear. For now I can at best assume that clarification is possible, point to a promising start in the attempt to do so—R. N. Giere, “Objective, Single-Case Probabilities and the Foundations of Statistics” ([2])—, and admit that the D-N-P model is viable only if sense can be made of propensities, or of objective, physical, lawful, single-case probabilities by any other name.

As for the requirement that explanations elucidate mechanisms, I can only repeat that an essential role is played in D-N-P explanations by the theoretical deduction of the probabilistic law(s) covering the explanandum.

In lieu of further exposition, I offer the beginnings of a defense, hoping thereby to sketch out the account a bit more fully in those areas most likely to be controversial.

BECAUSE IT APPLIES ONLY TO GENUINELY INDETERMINISTIC PROCESSES, OF WHICH THERE ARE FEW (IF ANY), D-N-P EXPLANATION IS TOO RESTRICTED IN SCOPE.

It is widely believed that the probabilities associated with standard gambling devices, classical thermodynamics, actuarial tables, weather forecasting, etc., arise not from any underlying physical indeterminism, but from an unknown or uncontrolled scatter of initial conditions. If this is right, then D-N-P explanation would be inapplicable to these phenomena even though they are among the most familiar objects of probabilistic explanation. I do not, however, find this troublesome: if something does not happen by chance, it cannot be explained by chance. The use of epistemic or statistical probabilities in connection with such phenomena unquestionably has instrumental value, and should not be given up. What must be given up is the idea that *explanations* can be based on probabilities that have no role in bringing the world's explananda about, but serve only

to describe deterministic phenomena.⁸ Whether there *are* any probabilities that enter into the mechanisms of nature is still debated, but the successes of the quantum-mechanical formalism, and the existence of “no hidden variable” results for it, place the burden of proof on those who would insist that physical chance is an illusion.

It could be objected more justly that D-N-P explanation is too broad, not too narrow, in scope. Once restrictions have been lifted from the value a chance may have in probabilistic explanation, virtually all explanations of particular facts must become probabilistic. All but the most basic regularities of the universe stand forever in peril of being interrupted or upset by intrusion of the effects of random processes. It might seem a fine explanation for a light’s going out that we opened the only circuit connecting it with an electrical power source, but an element of chance was involved: had enough atoms in the vicinity of the light undergone spontaneous beta-decay at the right moment, the electrons emitted could have kept it glowing. The success of a social revolution might appear to be explained by its overwhelming popular support, but this is to overlook the revolutionaries’ luck: if all the naturally unstable nuclides on earth had commenced spontaneous nuclear fission in rapid succession, the triumph of the people would never have come to pass.

No doubt this proliferation of probabilistic explanations is counter-intuitive, but contemporary science will not let us get away with any other sort of explanation in these cases—it simply cannot supply the requisite non-probabilistic laws. Because they figure in the way things *work*, tiny probabilities appropriately figure in explanations of the way things *are*, even though they scarcely ever show up in the way things turn out.

THE D-N-P MODEL BREAKS THE LINK BETWEEN PREDICTION AND EXPLANATION.

Hempel has justified a “qualified thesis of the structural identity of explanation and prediction” with this principle:

Any rationally acceptable answer to the question “Why did X occur?” must offer information which shows that X was to be expected—if not definitely, then at least with reasonable certainty. ([3], pp. 367–368)

Abundantly many D-N-P explanations—all those covering less than highly probable facts—violate this condition.

However, to abide by this condition and renounce explanations with meager probabilities I take to be worse. Why forgo the explanations of improbable phenomena offered by our theories, when these explanations provide as much of an account of why (and how) their explananda occur as do the explanations of “reasonably certain” phenomena that Hempel’s condition sanctions?

Too restrictive as it stands, Hempel's condition may be taken in a way not incompatible with D-N-P explanation. A D-N-P explanation does yield one prediction that is perfectly strict, to the effect that a certain physical probability exists in the circumstances given. If this probability fails to obtain, or to have the value attributed to it, the explanation must be false. It is a complaint against the world, not against the D-N-P model, that a direct, non-statistical test for the presence or value of this probability may prove impossible. Remarkably, the mechanisms of the world leave room for spontaneous nuclear disintegrations. Equally remarkably, our physical theory gives us insight into how they come about, and assigns determinate probabilities to them. These probabilities are connected to the rest of our theory by laws that permit both prediction and (where means exist) control: if undisturbed, nucleus *a* will have probability *p* to alpha-decay (so we should expect *a*'s decay with *epistemic* probability *p*); and if we wish to alter *p*, our theory tells us how *a* must be disturbed.

It has been objected to the view of probability taken in this paper that unless probability attributions are interpreted as predictions about how relative frequencies will *actually* come out in the long run, probabilistic laws lack empirical content. Thus if the relative frequency of decayed atoms in a large sample of some radioelement were, over a great length of time, to diverge significantly from the probability theoretically attributed to decay, that attribution would not be "borne out," that is, would be falsified. Otherwise, it is argued, probabilistic laws are compatible with all frequencies, and empirically vacuous.

But it is impossible for a world to "bear out" all of its probability-attributing laws in this sense. For these laws imply, among other things, that it is extremely unlikely that *all* actual long-run sequences will show a relative frequency near to the single-case probability. Therefore, the demand that all long-run decay rates nearly match all corresponding decay constants comes to a demand that nothing improbable show up in the long run, which is itself an improbability showing up in the long run. Intended to clear things up on the epistemological front, this proposal cannot even get out of its own way.

BY SPLITTING APART PROBABILISTIC EXPLANATION AND INDUCTION, THE D-N-P MODEL HAS LOST THE POINT OF PROBABILISTIC EXPLANATION.

Behind this objection lies the view that probabilistic (or statistical) explanation is an activity fundamentally unlike D-N explanation. A probabilistic explanation is seen as a piece of detective work. Unable to give a causal demonstration of the explanandum from evidence thus far assembled, we develop hypotheses, which are judged by how probable they are on the evidence, and whether they make the explanandum sufficiently probable. In the end, we put forward the most convincing inductive argument yet

found—the one making the explanandum most antecedently probable, given what else we know about events leading up to it.

This view of probabilistic explanation confuses epistemic with objective probability, and induction with explanation. Perhaps responsible for this confusion is the similarity of the tasks of explaining a phenomenon, gathering support for such an explanation, and gathering before- or after-the-fact evidence for a phenomenon's occurrence. This confusion is abetted by misleading ways of talking about "strong" or "good" explanations. We should distinguish the following. (i) A *strong (good) explanation* is one that has great theoretical power, regardless of how well-confirmed it is or how probable it holds the explanandum to be. (ii) A *strong (good) candidate for explanation* is a proffered explanation with well-confirmed *premises*, regardless of how probable it holds the explanandum to be and irrespective of how theoretically powerful it happens to be. (iii) A *strong (good) reason for believing that the explanandum fact will obtain* is furnished by before-the-fact evidence that leads, via one's theory, to an expectation of the explanandum with high epistemic probability. (iv) A *strong (good) reason for believing that the explanandum fact obtained* is given by any evidence that lends high epistemic probability to the proposition that the explanandum fact is a fact. Strong after-the-fact evidence, even for very improbable events, may be easy to come by. Reasons of types (iii) and (iv) need have nothing to do with explanation, and may be based on symptoms (Will it rain today?—Harry's rheumatism is acting up) or even less causally relevant information (Was Sue upset?—Her brother is certain she would have been).

Although the link between probabilistic explanation and induction is looser on the D-N-P model than on the I-S model, this is no fault: on Hempel's account it was entirely too close. Measuring the strength or "acceptability" of an explanation by the magnitude of the probability it confers on the explanandum blurs the distinctions just made. Keeping (i)–(iv) distinct, the D-N-P model enables us to state quite simply the object of induction in explanation: given a particular fact, to find, and gather evidence for, an explanans that subsumes it; given a generalization, to find, and gather evidence for, a higher-level explanans that subsumes it; in all cases, then, to discover and establish a true and relevant explanans. The issue of showing the explanandum to have high (relative or absolute) probability is a red herring, distracting attention from the real issue: the truth or falsity, and applicability, of the laws and facts adduced in explanatory accounts.⁹

■ | Notes

1. Let us say rather loosely that a system is deterministic if, for any one instant, its state is physically compatible with only one (not necessarily different) state at each other instant. A system is indeterministic otherwise, but lawfully so if a com-

plete description of its state at some one instant plus all true laws together entail a distribution of probabilities over possible states at later times

2 Although there is some difficulty in reconciling all that is said in [3] with this conclusion Hempel now accepts it (personal communication)

3 See, for example, [4]

4 A typographical error has been corrected

5 Causal relevance is established here via the wave equation. I do not mean to suggest that *causal* relevance is the only explanatory kind, cf. the mention of structural laws in section 1.

Some such notion of causal relevance appears to lie behind Salmon's "statistical-relevance" model of probabilistic explanation. Yet what matters is whether a factor enters into the probabilities present, not the statistics they produce.

6 Cf. the discussion of place selections and homogeneity in [6], sections 4 and 7. Salmon's criterion, which requires formal randomness, would here fail to distinguish a *randomly-produced* regular sequence from a *deterministically-produced* one. Notwithstanding formal similarities, only the latter is appropriately explained non-probabilistically.

7 Explanation (6) is true, however, only under the contrary-to-fact assumption—made for the sake of the example—of the existence of a class $-P$.

8 Of course, we might speak of statistical or epistemic probabilities as causes of, for example, beliefs. But if belief formation is not *physically* probabilistic, then probabilistic explanation of it would be impossible, in spite of this sort of causal involvement on the part of statistical or epistemic probabilities.

9 I would like to thank C. G. Hempel, Richard C. Jeffrey, and David Lewis for helpful criticisms of earlier drafts. I am especially indebted to David Lewis for the idea that a propensity interpretation of probability sits best with the account of probabilistic explanation given here. I have greatly benefited from discussions of related matters with Sam Scheffler and David Fair.

■ | References

- [1] Evans, R. D., *The Atomic Nucleus*. New York: McGraw Hill, 1965.
- [2] Giere, R. N., "Objective Single-Case Probabilities and the Foundations of Statistics." In *Logic, Methodology and Philosophy of Science*, vol. IV. Edited by P. Suppes, et al. Amsterdam: North-Holland, 1973.
- [3] Hempel, C. G., "Aspects of Scientific Explanation." In *Aspects of Scientific Explanation and Other Essays*. New York: Free Press, 1965.
- [4] Hempel, C. G., "Maximal Specificity and Lawlikeness in Probabilistic Explanation." *Philosophy of Science* 35 (1968), 116–33.
- [5] Jeffrey, R. C., "Statistical Explanation vs. Statistical Inference." In *Essays in Honor of C. G. Hempel*. Edited by N. Rescher et al. Dordrecht: D. Reidel, 1970.
- [6] Salmon, W. C., "Statistical Explanation." In *Statistical Explanation and Statistical Relevance*. Edited by W. Salmon. Pittsburgh: University of Pittsburgh Press, 1971.

- 6.1 Hempel's D-N Model 767**
Carnap on the Motivation for Hempel's D-N Model 767
Hempel's Conditions of Adequacy for D-N Explanation 769
Elliptically Formulated Explanations 772
Partial Explanations and Explanation Sketches 773
- 6.2 Hempel's D-N Model and the Thesis of Structural Identity 773**
The Syphilitic Mayor 775
Evolutionary Theory 775
Scriven's Bridge 777
Koplik Spots 778
The Flagpole and the Pendulum 778
- 6.3 Hempel's I-S Model 779**
Hempel's Conditions of Adequacy for I-S Explanation 779
The Problem of Ambiguity in Statistical Explanation 781
Hempel's Requirement of Maximal Specificity (RMS) 783
- 6.4 Ruben on the Irrelevance Objection to Hempel's Models of Explanation 784**
The Arsenic Eater 785
The Birth-Control Pills 785
The Hexed Salt 786
A Proposed Cure for the Irrelevance Problem: The Causal Condition 787
- 6.5 Ruben's Single-Statement View of Explanation 788**
Ruben on the Role of Laws in Explanation 789
- 6.6 Railton on What Is Wrong with Hempel's Models of Explanation 790**
Nomic Subsumption 790
Inductive Arguments 791
The High-Probability Requirement 792
The Maximal Specificity Requirement 792
- 6.7 Railton's Deductive-Nomological Model of Probabilistic Explanation 793**
Railton's Conditions of Adequacy for D-N-P Explanation 794
Propensities and Probabilities 796
Railton's Response to Some Objections 797
- 6.8 Summary 799**

6 | COMMENTARY

6.1 | Hempel's D-N Model

The phrase *deductive-nomological* is an apt description of the model that Hempel first proposed in his classic paper published in 1948.¹ For according to Hempel, many scientific explanations are deductively valid arguments having at least one statement of an empirical law in their premises. (The adjective *nomological* is derived from the Greek word *nomos*, meaning "law.") Only later, in 1962, did Hempel focus his attention on statistical explanation.² He then proposed his I-S model, according to which statistical explanations are inductive arguments with at least one statement of an empirical statistical law in their premises. Thus, both models are instances of the covering law thesis that all explanations are arguments (either deductively valid or inductively strong) that must involve laws. Before discussing Hempel's models in detail, it will be helpful to consider the underlying motivation for Hempel's proposals.

CARNAP ON THE MOTIVATION FOR HEMPEL'S D-N MODEL

As Rudolf Carnap points out in "The Value of Laws: Explanation and Prediction," scientists and philosophers have not always valued theories for their explanatory power. In the second half of the nineteenth century, for example, philosopher-scientists such as Ernst Mach and Pierre Duhem insisted that the proper function of scientific theories was not to explain phenomena but merely to classify and summarize experimental laws.³ Duhem defines explanation as follows: "To explain (explicate, *explicare*) is to strip reality of the appearances covering it like a veil, in order to see the bare reality itself."⁴ Duhem then argues that if scientific theories are intended to explain (in his own sense of stripping reality bare), then they will inevitably make science subordinate to metaphysics. The kinetic theory of gases and the wave theory of light, for example, posit the existence of atoms and the optical aether as an essential part of the explanations they provide. But neither atoms nor the aether can be observed directly. At best claims about these "bare realities" can be tested only indirectly by seeing what experimental laws they entail. Moreover, these experimental laws are compatible with a wide range of different assumptions about atoms and the aether. Thus, testing will be unable to confirm any particular theory (or the unobservable entities it posits) relative to its rivals. So, Duhem concludes, the assertion that atoms exist or that there really is an aether are metaphysical claims that empirical science can neither confirm

nor refute.⁵ Explanation is not, and cannot be, one of the aims of scientific theories.

Carnap rejects the claim of Duhem, Mach, and other positivists that science cannot explain; nevertheless he is sympathetic to their skepticism about theories involving metaphysical assumptions about unobservable entities and causes. Carnap insists that scientific theories can give genuine explanations but only if the explanations involve testable, empirical laws. Consider the example Carnap discusses of the German biologist and philosopher Hans Driesch (1867–1941). Driesch, who had done pioneering work on embryology and limb regeneration in sea urchins, discovered in 1891 that prior to the fifth division of a fertilized sea-urchin's egg any cell was capable of developing into a complete embryo. These and other findings led Driesch to espouse a "vitalistic" account, according to which an inner force or purpose in each living organism—what Driesch called an *entelechy*—is responsible for maintaining the integrity of the organism, directing development, and regenerating lost parts. In a number of works, Driesch championed his entelechy theory as an explanation for biological and psychological phenomena. He saw an important similarity between an embryo developing into an adult organism and a person voluntarily deciding to perform one action rather than another. In both cases, he argued, we cannot predict what will happen on the basis of the laws of physics and chemistry. We can explain what has happened after the event, but only because we then know how things have turned out. Driesch regarded the power of a human being to choose which action to perform as a refined manifestation of the teleological, directive power of the entelechy that is in every living creature. And what is an entelechy? It is a nonphysical, nonmaterial, nonspatial inner force that directs everything that an organism does, from cellular processes to voluntary actions. Every living thing has an entelechy, each species has its own distinctive kind, and it is possession of an entelechy that distinguishes living things from machines.

Despite their respect for Driesch's scientific work and his genuine desire that philosophy take science seriously, Carnap and Reichenbach could not accept that Driesch's entelechy theory really explained anything. In defending the entelechy theory against this accusation, Carnap recalls Driesch retorting that his introduction of the term *entelechy* to explain the behavior of organisms was no different from physicists introducing the term *magnetism* to explain the behavior of magnets and bits of iron. After all, we can neither see nor touch the force of magnetism. All we can actually observe is the motion of bodies, which magnetism was posited to explain. Similarly, Driesch maintained, it is legitimate for him to introduce entelechies to explain biological phenomena. Carnap responded by pointing out that the cases are relevantly different. For when physicists introduced the term *magnetism*, they did not simply posit the existence of

an unobservable entity; they also specified laws that magnetized bodies must obey. These laws can be used to make predictions that can be tested by experiment and observation. Driesch's entelechy theory specifies no such laws and is thus completely lacking in predictive power. Therefore, Carnap concludes, Driesch's theory does not give genuine explanations.

Our discussion of Carnap's criticisms of Driesch's entelechy theory brings to light one of the central motivations for Hempel's D-N model of explanation, namely, the requirement that every genuine scientific explanation include at least one empirical law in its explanans. Another closely related theme is Carnap and Hempel's insistence that explanation and prediction go hand in hand: every genuine explanation must be capable of predicting its explanandum. Later we will focus our attention on Hempel's defense of what he calls the *thesis of structural identity*, that, in all formal respects, explanations are the same as predictions, and predictions the same as explanations.

HEMPEL'S CONDITIONS OF ADEQUACY FOR D-N EXPLANATION

The requirement that the explanans include at least one empirical law is obviously not sufficient for an explanation for two reasons. First, just including a law, any law, is not enough; the law must be essential to the derivation of the explanandum. Clearly, we would not be much impressed by a purported explanation in which the laws mentioned were entirely irrelevant to the event or phenomenon needing explanation. Second, all by themselves laws do not entail that any specific thing will happen. Thus, when we seek to explain the occurrence of an event, we must also include in the explanans statements of various initial conditions that, in conjunction with the laws, logically imply the explanandum.

We thus arrive at the general scheme of Hempel's D-N model of explanation, which he summarizes in the section "Two Basic Types of Scientific Explanation":

C_1, C_2, \dots, C_k	Statements of particular facts and initial conditions	}	Explanans
L_1, L_2, \dots, L_r	General laws		
<hr/>			
E	Description of the event, law, or fact to be explained		Explanandum

In his 1948 article, Hempel gave the criteria for a D-N explanation in the form of four conditions of adequacy, which he divided into two groups: logical and empirical.

■ LOGICAL CONDITIONS OF ADEQUACY FOR D-N EXPLANATION

- R1 The explanandum must be a logical consequence of the explanans
- R2 The explanans must contain general laws, and these must be essential for the derivation of the explanandum.
- R3 The explanans must have empirical content; that is, it must be capable, at least in principle, of test by experiment or observation.

■ EMPIRICAL CONDITION OF ADEQUACY FOR D-N EXPLANATION

- R4 The sentences in the explanans must be true

Before discussing Hempel's defense of the thesis of structural identity (that is, his thesis that explanation and prediction are formally identical) and his other model of explanation (the I-S model), a few brief remarks about Hempel's criteria of adequacy for a D-N explanation will be helpful in understanding criticisms of his views. We offer brief comments on each criterion in turn.

Why does Hempel require that the explanans logically entail the explanandum? The reason is simple. If we have genuinely explained why the explanandum event has occurred, then we must have given sufficient grounds for expecting that the event in question would occur. After all, the question "Why did E_1 occur?" usually arises when other outcomes seem possible: why did E_1 occur, rather than, say, E_2 or E_3 ? A completely satisfactory answer to this question would thus show that E_1 was the only event among the possible alternatives that could have occurred. Ideally, then, an explanation must be a logically valid argument, for in a valid argument, if the premises are true, then the conclusion also has to be true. The best kind of explanation is one in which, given the information in the explanans, the explanandum has to be true. If it were still possible, given the explanans, for the explanandum to be false, then we would not have explained, fully and completely, why the explanandum event occurred. As we shall see later in this commentary, Hempel is forced to relax this deductive standard of explanation in order to accommodate statistical explanations of particular events.

Our discussion of Carnap's criticisms of Driesch has already touched on the motivation for requiring that every explanation contain at least one law. Two additional features of this requirement are noteworthy. First, Hempel does not require that the laws be causal. As far as Hempel is concerned, there can be perfectly satisfactory scientific explanations solely in terms of Snell's law, Hooke's law, the equation of state for an ideal gas, or the like. These laws, expressed in the form of equations, are sometimes called *functional laws* because they specify the mathematical function re-

lating the value of one variable (such as pressure) to the value of other variables (such as temperature and volume). (For more on functional laws, see the section "The 'Missing Values' Problem for Functional Laws" in the commentary on chapter 7.) Thus, Hempel deliberately does not limit scientific explanation to causal explanation. A second noteworthy feature of (R2) and the other logical conditions of adequacy is that Hempel does not require that the explanans contain any statements of initial conditions. Of course, such initial conditions must be included in any deductive explanation of the occurrence of a particular event, but Hempel wants his model to cover not only the explanation of events but also the explanation of laws. That is, he wants his model to apply to cases in which one law or a set of general laws (such as Newton's laws of motion and gravitational attraction) explain another law (such as Kepler's second law) by deductively implying it. (The explanation of one law by another, more general, law or theory is discussed in chapter 8. In the present chapter, the focus is on the explanation of events.)

Condition (R2) mentions only general laws, which would include laws of mathematics. While laws of mathematics often play an important role in scientific explanations, Hempel requires that at least one of the laws in the explanans be empirical. It is perfectly legitimate for the explanans to include mathematical laws (such as the Pythagorean theorem or the principles of algebra), but it must also include at least one empirical law that, unlike the laws of mathematics, can be tested by observation or experiment. Strictly speaking, (R3) is redundant, since it follows from (R1): anything that logically implies the explanandum, which has empirical content, must itself have empirical content. But such is the importance of (R3) to the spirit of Hempel's conception of explanation that it is listed as a separate requirement.

The three conditions (R1), (R2), and (R3) are grouped together under the heading of logical conditions of adequacy in Hempel's 1948 article because he thinks that we can tell whether something satisfies them without our needing any empirical information about the world.⁶ But the last condition, (R4), is different. This condition requires the sentences in the explanans to be true. But for most if not all of these sentences their truth or falsity can only be determined empirically.⁷ Elsewhere, Hempel calls a set of sentences that satisfies the logical conditions of adequacy a *potential explanans*. It is only when those sentences are actually true that we have a genuine explanation. As far as the D-N model is concerned, Hempel's notion of a correct or genuine explanation is strongly objective. A group of scientists may believe that they have given an explanation, and they may indeed have considerable justification for their belief, but unless the potential explanans they provide is actually true, no genuine explanation has been given. Thus, for example, Hempel would deny that the phlogiston theory of chemistry explained why metals burn in air (calcination). Why? Because the phlogiston theory is, as a matter of fact, false. Of course,

we can say that phlogiston theory explained the calcination of metals, and this is often the simplest way of expressing ourselves. But if we accept Hempel's empirical truth condition (R4) for explanation, then what we really mean is that the phlogiston theory would have explained the burning of metals in air if it had been true

ELLIPTICALLY FORMULATED EXPLANATIONS

Hempel recognizes that the explanations scientists actually give often fall short of the conditions of adequacy for D-N explanation. But Hempel does not see this failure of his model to describe actual scientific practice as in any way refuting the D-N model as an account of scientific explanation. Partly this is because Hempel's model is intended to be normative rather than merely descriptive: his aim is to articulate an ideal of what a good scientific explanation *should* be like, not simply to summarize or describe the explanations that scientists actually give. Hempel arrives at his model not by ignoring scientific practice, but by reflecting on paradigmatic examples of scientific explanation (such as those mentioned in his articles) and then isolating their essential features. With his D-N model clearly formulated, Hempel thinks he can explain why many actual scientific explanations fail to satisfy his model by appealing to pragmatic factors.

In many respects, Hempel's attempt to clarify the concept of explanation is similar to the analysis of the concept of proof given by logicians and philosophers of mathematics. The actual proofs that mathematicians write down often leave out steps (considered obvious or trivial), but this in no way undermines or refutes the strict, formal concept of proof that serves as our ideal.

The most common way in which the explanations given by scientists deviate from Hempel's D-N model is by being incomplete. For example, why does ice float on water? Most scientists would accept as a satisfactory explanation the assertion that, unlike most substances, water expands when it freezes. This is an instance of what Hempel calls an *elliptically formulated explanation*. Some laws or facts are left out of the explanans of such explanatory arguments. Why? Because the relevant laws and facts (such as Archimedes' principle of buoyancy and the inverse relation between density and volume) are so well known and accepted (at least by other scientists) that to write all of them down would be tedious and a waste of time. Consequently, the explanandum does not actually follow logically from the explanans, but the missing material can easily be filled in, and once it is filled in, the amended explanans does logically imply the explanandum. So this kind of deviation from the D-N model is innocuous and justified on pragmatic grounds. Far from refuting the D-N model, such examples actually confirm it, since the D-N account is able to explain in a plausible way why so many actual explanations are incomplete, and in making them more complete we come closer to the D-N ideal.

PARTIAL EXPLANATIONS AND EXPLANATION SKETCHES

There are other, more radical, kinds of explanatory incompleteness discussed by Hempel, namely *partial explanations* and *explanation sketches*. In a partial explanation, even when all the missing premises are added to the explanans, it still does not logically imply the explanandum. What does follow from the supplemented explanans is not the original explanandum but something more general. For example, we might seek to explain why some kinds of apple turn from green to red as they ripen. Suppose that we give an explanation but, even when our explanans is filled out, what it actually entails is that apples change color as they ripen. Since the change from green to red in some kinds of apple is a special case of this more general sort of change, to that extent, Hempel is prepared to call such explanations "partial." But strictly speaking, a partial explanation does not actually explain the explanandum it was invoked to explain. Hempel offers a similar example from the *Psychopathology of Everyday Life* in which Freud attempts to explain why he wrote the wrong date in his diary. What Freud actually explained (even when his explanans is filled out with the psychological facts and the presumed laws of Freudian psychology that are missing from his account) is not the particular slip he made about the date, but why he performed an action that, in some way or other, symbolically represented the fulfillment of one of his subconscious wishes.

The final category of incomplete explanations, *explanation sketches*, consists of those explanatory accounts that are so vague and incomplete—in short, so sketchy—that they fail to qualify as either elliptical or partial explanations. At best, an explanation sketch provides a general outline that might prove capable of being developed into a satisfactory explanation at some future time.

6.2 | Hempel's D-N Model and the Thesis of Structural Identity

In "The Thesis of Structural Identity," Hempel formulates his thesis as the conjunction of two claims or subtheses:

every adequate explanation is potentially a prediction; and
every adequate prediction is potentially an explanation.

Because Hempel's thesis asserts that there is a symmetry between explanation and prediction, his thesis of structural identity is often called the *symmetry thesis*, and criticisms of the thesis are instances of the *symmetry*

objection. In his article, Hempel attempts to defend both parts of his thesis of structural identity against alleged counterexamples.

In order to understand Hempel's defense, several features of his thesis need to be underscored. First, although many instances of the symmetry objection focus on deductive explanations, Hempel believes that his thesis applies to both D-N and I-S explanations.

Second, Hempel uses the term *prediction* in a special sense. Ordinarily, we would say that Jones has predicted the outcome of the Super Bowl if he announces the name of the winning team before the game is played. In this sense, a prediction is simply a statement that need not be accompanied by any supporting reasons or argument. But since Hempel is primarily interested in the explanations and predictions given by scientists, he deliberately restricts the meaning of the term *prediction* to *predictive argument*. Without this restriction, his thesis would be obviously false.

Third, Hempel intends his models of explanation to cover both the explanation of laws and the explanation of events. But it makes little sense to talk about the prediction of laws, since laws are not the sort of things that happen at any particular time. So Hempel implicitly restricts his symmetry thesis to the explanation and prediction of events.

Finally, Hempel recognizes that we can sometimes use laws and theories to make inferences about what has happened in the past on the basis of initial conditions and particular facts that hold at a later time. These inferences are often called *postdictions* or *retrodictions*, and scientists ordinarily regard them as a species of prediction, for just like predictions about the outcome of an experiment, postdictions involve the deduction from a law or theory of something whose truth is not yet known. But Hempel is careful to exclude postdictions from the scope of his symmetry thesis. The reason is obvious: even though we can use Newtonian mechanics to infer from the present positions of the sun, the moon, and the earth that a total solar eclipse occurred two thousand years ago, we cannot accept that the way things are now explains the way things were two thousand years ago; present events cannot explain past events. So the symmetry thesis is concerned solely with predictive arguments in Hempel's restricted sense, that is, with arguments in which all the initial conditions and particular facts hold at times prior to the event described in the conclusion.

With these preliminaries in place, we may now consider objections to the symmetry thesis. Counterexamples to the first subthesis must be adequate explanations that are not potentially predictions; counterexamples to the second subthesis must be adequate predictions that are not potentially explanations. We begin with three objections to the first subthesis—alleged examples of adequate explanations that could not have been used to predict the events that they explain. For convenience, we shall assign them titles.

THE SYPHILITIC MAYOR

Michael Scriven asks us to consider a certain medical patient—the mayor of his town, let us suppose—whose name is Jones. Consider the question, Why did the mayor, Jones, contract paresis? Paresis is a form of general paralysis that affects only those who have had untreated syphilis for many years. This being so, the answer to our question is, Because Jones had untreated syphilis, and the only cause of paresis is syphilis. This, allegedly, is an explanation of why Jones now has paresis. But since only 10 percent of untreated syphilitics go on to develop paresis, Jones's syphilis could not have been used to predict his paresis. Indeed, since 90 percent of syphilitics do not contract paresis, we would have predicted exactly the opposite.

Hempel's reply to this objection is short and sweet: no adequate explanation has been given. Merely to cite one condition that is necessary (but not sufficient) for the occurrence of an event—even if that condition is based on a law—is not to explain that event. Presumably, it is a law that all people who die have needed to breathe oxygen while they were alive. But we do not suppose that we have explained why someone has died merely by pointing out that the person was an oxygen breather. Similarly, to use one of Hempel's own examples, we do not think that we have explained why a particular person won the Irish sweepstakes simply by adducing the fact that the person had bought a ticket.

Many people have judged Hempel's reply to be unsatisfactory, in part because, in the syphilitic mayor example, we know of no other factors that can influence the chance of someone's being afflicted with paresis. There is thus an understandable tendency to regard this example as an event of low probability being explained by a statistical law. So what may in fact be at issue here is whether we should insist, as Hempel does, that any statistical explanation confers high probability on its explanandum. Hempel's high-probability requirement will be discussed later when we examine the I-S model and the criticisms of it.

EVOLUTIONARY THEORY

Several versions of the evolutionary theory objection have been raised by Hempel's critics. In its simplest form, the objection runs as follows:

Darwin explained the origin of species using this theory of natural selection working on random variations. Scientists accept that Darwin's theory offers genuine explanations, yet no scientist has been able to use Darwin's theory to predict the coming-into-existence of any new species. Thus, evolutionary theory explains but it does not predict.

In reply, Hempel stresses the importance of distinguishing between what he calls the *story* of evolution and the *theory* of evolution. The story

of evolution is a narrative describing the sequence of species that have arisen and become extinct since life first appeared on earth. Even if this narrative is completely true, it has no explanatory import whatever. It is merely a description of what has happened in the past. The theory of evolution, by contrast, employs generalizations about heredity, mutation, and selection plus a host of detailed assumptions about environmental conditions and ecological relations. At best, this theory can offer only partial, probabilistic explanations of general facts about species survival and extinction. What it cannot do, at least in the present incomplete state of our biological knowledge, is explain why any particular species came into existence when it did. In short, by arguing that evolutionary theory explains considerably less than one might have supposed, Hempel denies that evolutionary theory explains what it is unable to predict. Biologists may have reasonably good explanations of why, following the extinction of the dinosaurs, other species were able to flourish and evolve, but they cannot yet explain why, say, a particular species of rat or aardvark evolved when it did, with the characteristics it did. The random nature of the variations on which natural selection works precludes explaining or predicting anything very detailed about the coming-to-be of new species. Like the syphilitic-mayor objection, then, this one fails because it is based on a false assumption (about the supposed extent of explanation offered by evolutionary theory).

A closely related objection to Hempel's theory, also involving evolutionary theory, was raised by Michael Scriven.⁸ Even after a new trait has appeared in an organism, it is difficult to know whether that trait is adaptive. In particular, it is hard to judge the magnitude of the advantage, if any, that a new trait (such as a larger shell) confers on its possessor. Moreover, there is an ineliminable element of chance involved in determining which individuals will actually survive. Even the fittest giraffe might be killed by lightning before having the opportunity to mate and produce offspring. But once natural selection has operated for some time on many thousands or millions of individual organisms, it is much easier to identify which traits are adaptive. Clearly, we must confront the threat of a circularity of definition here—namely, that of simply identifying the fittest organisms with those that actually survive.⁹ But that aside, we must grant that sometimes the information needed to explain an event can be obtained only by making inferences from the fact that the event in question has actually occurred. Hempel refers to explanations having this feature as *self-evidencing* explanations. This notion plays a major role in Hempel's treatment of the next objection, also proposed by Scriven, of a collapsing bridge.

SCRIVEN'S BRIDGE

Sometimes the only ground we have for asserting that some statement in the explanans is true lies in our knowledge that the explanandum event did, in fact, occur. This feature is at work in Scriven's example of the collapsed bridge. The collapse tells us not only that metal fatigue occurred, but that it was serious enough to cause the failure of the entire structure. Similarly with the man who kills his wife out of jealousy or the patient who develops skin cancer after exposure to ultraviolet light: in all such cases, we could not have predicted the relevant events, but we can nevertheless explain them after they occur. Here again, contrary to the symmetry thesis, we seem to have explanations that are not predictions.

Hempel agrees that Scriven's bridge is a case in which we would not have had all the information necessary for predicting the collapse prior to its occurrence. But, Hempel insists, this does not mean that Scriven has given a counterexample to the first subthesis because, when interpreted correctly, that subthesis makes merely a *conditional* claim, namely:

if all the information in the explanans had been known and taken into account before the occurrence of the explanandum event, then the event could have been predicted.

What Scriven has done is to show that, in some cases, the antecedent of this conditional is not satisfied. But this does not show that the conditional is false.

Hempel introduces the term *self-evidencing* to describe those explanations in which the information that the explanandum statement, *E*, is true provides a crucial evidential support for one of the particular statements in the explanans, *C*₁. Hempel insists that such explanations are not circular, since they are not being used to establish that *E* is true. As with any explanation, we already know (or presume that we know) that *E* is true, that the explanandum event happened as described. Thus, although we are using *E* as part (or even the whole) of the evidence for *C*₁, we are not then using *C*₁ as evidence for *E*. Hence, there is no epistemic circularity. Moreover, although we are using *C*₁ to explain *E*, we are not also using *E* to explain *C*₁. Hence, there is no explanatory circularity.

Hempel concedes that the second subthesis of the symmetry thesis is less secure than the first. That is, Hempel admits that there seem to be adequate predictions that are not potentially explanations. An example of his own invention, involving the association between Koplik spots and the measles, serves to illustrate the problem.

KOPLIK SPOTS

Koplik spots are tiny, whitish spots that appear on the inside of one's cheeks about a week before one succumbs to a full-blown case of measles. Supposing that the appearance of Koplik spots is *always* followed by the measles, the connection between them may be judged lawlike. Such a law can then be used to predict that a patient with Koplik spots will have measles a week later. Still, the Koplik spots do not explain why the patient will develop full-blown measles in a week's time.

Hempel suggests that our reluctance to regard the Koplik-spots argument as explanatory likely reflects our doubts about whether measles do in fact always follow the spots as a matter of universal law. Perhaps, he conjectures, we could produce Koplik spots by injecting a small quantity of the measles virus into someone's cheek without the spots then being followed by full-blown measles. But this response to the Koplik spots case is not entirely satisfactory. For even if the relation between Koplik spots and the measles is not one of universal law but merely one of high probability, it remains unclear why the resulting argument would not satisfy the conditions of Hempel's I-S model and hence qualify as a statistical explanation (Hempel's I-S model is discussed in section 6.3 of this commentary.)

To some of his critics, Hempel's models of explanation seem vulnerable to the Koplik spots example because the models include no condition that mentions causation. Why do the Koplik spots fail to explain the later case of full-blown measles? Because the spots are not the cause of the measles. Rather, the spots and the full-blown measles are both joint effects of a common cause, namely, infection with the measles virus. In just the same way we can use the falling reading on a barometer to predict that a storm is approaching, but we do not take the barometer reading to be the cause of the storm's approach, nor do we take it to be an explanation of its approach. Again, we have prediction without explanation. The falling reading on the barometer and the approach of the storm are joint effects of a common cause, namely, a drop in atmospheric pressure. (In "Arguments, Laws, and Explanation," David-Hillel Ruben examines whether Hempel's models of explanation can be rescued by requiring that the explanans contain a statement describing the cause of the explanandum event.)

THE FLAGPOLE AND THE PENDULUM

Although Hempel does not consider these particular examples, they are often cited by critics of the symmetry thesis. A flagpole of height H , casts a shadow of length S . Given the law that light travels in straight lines, and the elevation of the sun, θ , we can deduce the length of the shadow from the height of the flagpole. Thus, we have both an adequate prediction of

So and, let us assume, an explanation of why the shadow has that particular length by using the equation $H = S \tan \theta$. But given that same equation, we could have just as easily deduced the height of the flagpole from the length of its shadow. Although perfectly fine as a prediction, this would not be accepted as an explanation: the length of the shadow does not explain the flagpole's height. The pendulum example is similar. From the period of a simple pendulum (the time it takes to perform one complete oscillation) we can deduce (i.e., predict) the length of the pendulum, but we do not think that the pendulum's period explains its length. Thus, on the face of it, the flagpole and the pendulum seem to provide plausible counterexamples to the second subthesis: not every adequate prediction is potentially an explanation.

6.3 | Hempel's I-S Model

As we have seen, Hempel's D-N model construes explanations as deductive arguments. But Hempel was aware all along that some scientific explanations could not be reconstructed in this fashion. This is especially true of theories in physics and genetics that use probabilistic laws to explain particular events. In 1962, Hempel turned his attention to these other, nondeductive arguments and formulated his inductive-statistical (I-S) model of explanation.

HEMPEL'S CONDITIONS OF ADEQUACY FOR I-S EXPLANATION

Although by 1962 Hempel no longer required that all explanations deductively entail the events they explain, he continued to defend the covering law thesis: just like D-N explanations, I-S explanations must have at least one (in this case statistical) law among their premises. Hempel also retained a substantial part of the concept of explanation that motivated the D-N model. Like their D-N counterparts, I-S explanations are still arguments, and while the conclusions of I-S arguments no longer follow from their premises with logical necessity, Hempel insisted that the explanans must make the explanandum highly probable. The higher the probability, the stronger the argument and the better the explanation. The strength of an I-S explanation is measured by the inductive probability of its conclusion relative to its premises. Hempel assumed that this strength is equal to the numerical value of the probability given by the statistical law in the explanans. Thus, given the statistical law that *F*s are very likely to be (or be followed by) *G*s, and given some particular fact that *a* is *F*, it is very likely that *a* is *G*. An I-S explanation therefore has the following schematic form:

$P(G/F) = 0.95$	Statistical law	} Explanans
Fa	Particular fact	
<hr/>		
Ga	[0.95]	Explanandum

Note that the conclusion of this I-S argument is not "*a* is almost certain to be *G*" but the unqualified statement "*a* is *G*." It is the fact that *a* is *G* that this inductive argument purports to explain. Hempel insists that expressions such as "*a* is almost certain to be *G*" are incomplete and thus neither true nor false. All meaningful, empirical statements of probability must be qualified as being relative to some body of evidence. It is either true or false that *a* is or will be *G* what the full I-S explanation expresses is that the truth of *Ga* is very probable relative to the statistical law and the particular facts.

From everything we have said thus far, D-N explanations would seem to be a limiting case of I-S explanations. when the argument is deductively valid, the inductive probability of the conclusion relative to the premises is 1. Hempel denies this. He sees I-S explanations as essentially different from D-N explanations because, he insists, I-S explanations must be relativized to a particular "knowledge situation" (712). This epistemic relativity of I-S explanations arises from the requirement of maximal specificity (RMS), which Hempel imposes on all statistical explanations of particular events as a way of solving a certain problem of ambiguity that infects such explanations.¹⁰ Hempel's RMS, and the problem of ambiguity it is meant to solve, are discussed in the next section. For the present, we will summarize the conditions of adequacy for an I-S explanation of a particular fact or event.

■ LOGICAL CONDITIONS OF ADEQUACY FOR I-S EXPLANATION

- S1 The explanandum must follow from the explanans with high inductive probability.
- S2 The explanans must contain at least one statistical law, and this must be essential for the derivation of the explanandum.
- S3 The explanans must have empirical content; that is, it must be capable, at least in principle, of test by experiment or observation.

■ EMPIRICAL CONDITIONS OF ADEQUACY FOR I-S EXPLANATION

- S4 The sentences in the explanans must be true.
- S5 The statistical law in the explanans must satisfy the requirement of maximal specificity.

THE PROBLEM OF AMBIGUITY IN STATISTICAL EXPLANATION

As explained in his article, "Inductive-Statistical Explanation," Hempel proposed his RMS in response to the problem of ambiguity in statistical explanation. The problem of ambiguity can be regarded as an instance of a more general difficulty arising whenever we wish to use statistical information about classes of cases or events to decide the probability of a single case or event. This is the so-called problem of the single case. for any event that we wish to explain, there will be many different reference classes to which the event could be assigned; each choice of a reference class will present us with a different statistical law, and often these laws will have different probabilities associated with them. Let us consider one of Hempel's own examples.

We wish to explain why a particular day, n , November 27 in Stanford, has the property, W , of being warm and sunny. Thus the explanandum is Wn . Among the many reference classes to which n belongs is (on the one hand) the class, N , of November days in Stanford, and the probability of warm weather on such a day, $P(W/N)$, is 0.95. So if we assign n to the reference class N , the high-probability condition is met and, apparently, we have an I-S explanation of Wn . That is, we have explained why November 27 in Stanford was warm and sunny by pointing out that n belongs to class N and citing the statistical law that says that the probability of warm and sunny days in that class is very high. But November 26 in Stanford was cold and rainy, and so n also belongs (on the other hand) to a different reference class, S , of immediate successors of cold and rainy days in Stanford. Assume, with Hempel, that $P(W/S) = 0.2$. Thus, $P(\sim W/S) = 0.8$, which we may agree qualifies as high. Now if, contrary to fact, n had *not* been warm and sunny, then we could have used the law $P(\sim W/S) = 0.8$ to explain why November 27 in Stanford was *not* warm and sunny. This, to Hempel, is intolerable. Of course, since n was warm and sunny, we could not use $P(\sim W/S) = 0.2$ to explain Wn . But, nonetheless, without some further condition of adequacy for I-S explanations, we are in the position of being able to "explain" the weather on November 27 in Stanford either way, whether the day was warm and sunny or not. For Hempel, this possibility of "explaining" an event whether or not it occurred means that no genuine explanation has been given at all.

We can state the problem of ambiguity as follows: given an I-S explanation with true premises of some explanandum, Ga , there will often be another I-S explanation with true premises and conclusion, $\sim Ga$. For convenience, let us call these two arguments (1) and (2):

Argument 1

$$P(G/F) = 0.95$$

$$\frac{Fa}{Ga} \quad [0.95]$$

Argument 2

$$P(\sim G/H) = 0.96$$

$$\frac{Ha}{\sim Ga} \quad [0.96]$$

Now, it might be thought that ambiguity is not really a problem at all. Since an explanation is sought only once the explanandum event has occurred, we would never actually accept both (1) and (2) as correct explanations. But Hempel judges this way of dismissing the problem to be unsatisfactory. Why? Because it is still the case that the premises of both (1) and (2) are true, and, let us assume, both are contained in our body of knowledge, *K*. (More will be said about Hempel's notion of a body of knowledge presently.) So if *a* had turned out not to be *G*, we could have given an equally strong explanation for $\sim Ga$. In what sense, then, have we explained the fact that *Ga* if by appeal to truths in *K* we could just as well have explained the fact that $\sim Ga$? To put the point in what may be a clearer form: if the essence of explanation is nomic expectability (that is, predictability based on laws), then we cannot accept that (2) would be just as good an explanation as (1) if *a* had failed to be *G*. Clearly, there is a connection between Hempel's concerns over explanatory ambiguity and his commitment to the thesis of structural identity (the symmetry thesis) between predictions and explanations. *K* contains the premises of (1) and the premises of (2). Thus, if we are justified in predicting *Ga*, we would also be justified in predicting $\sim Ga$. The fact that *Ga* turns out to be true and $\sim Ga$ false cannot be given as a reason for judging one of these predictions justified and the other not. We cannot without further restrictions allow that both predictions are justified, for we would then be justified in accepting a contradiction. Although *K* remains consistent under logical implication, it would not remain consistent under unrestricted inductive inference. Arguments like (1) and (2) would inevitably generate what Hempel has elsewhere called "inductive inconsistencies."¹¹

We might try to avoid the problem of ambiguity by adopting Carnap's requirement of total evidence on all applications of inductive logic. This requirement demands that we use all the evidence available in determining degrees of confirmation (inductive probability), allowing us to use a part of the total evidence only if the evidence we ignore is irrelevant to the conclusion (i.e., only if the conclusion has the same probability given the relevant part of the evidence as it has given the total evidence). Hempel believes that this suggestion is on the right track. The task is to refine Carnap's requirement so that it addresses the specific problem of ambiguity in statistical explanation. As Hempel notes, we should not interpret Carnap's requirement of total evidence as demanding that we use all the information available to us. For in that case all probabilistic explanations offered at a given time would have the same (very large) explanans. More-

over, when we offer an explanation, we already know that the explanandum event has occurred. Including this fact in the premises would make the argument trivially deductively valid (not inductive) and nonexplanatory (since no law in the premises would be essential to the deduction of the conclusion). So we must limit Carnap's requirement of total evidence to just that evidence that is of potential explanatory relevance to the explanandum-event. In short, we must express the desired condition in such a way that I-S arguments satisfying it will use the right reference class for the purposes of explaining the explanandum-event. Here is Hempel's proposal.

HEMPEL'S REQUIREMENT OF MAXIMAL SPECIFICITY (RMS)

Consider our standard statistical explanation schema:

$$\frac{P(G/F) = r}{\frac{Fb}{Gb}} [r]$$

Let S be the conjunction of the premises, and let K be the total set of statements accepted at the time the explanation is proposed. Hempel's RMS stipulates that if $(S \& K)$ implies that b belongs to a class F_1 , and that F_1 is a subclass of F , then $(S \& K)$ must imply a statement specifying the probability of G in subclass F_1 , say $P(G/F_1) = r_1$. Here, r_1 must equal r unless the probability statement $P(G/F_1) = r_1$ is simply a theorem of mathematical probability theory.

As its name indicates, the RMS insists that we assign the explanandum event, Gb , to the most specific reference class (the maximally specific reference class) to which it is known to belong. To understand better how the requirement works, we may consider again the explanation of why the weather in Stanford on November 27 was warm and sunny and the problem of epistemic ambiguity arising from the rival arguments (1) and (2). (Recall that argument (1) begins with the probability that, relative to the class N of days in November, day n has the property W of being warm; argument (2) begins with the probability that, relative to the class S of days succeeding cold and rainy days in Stanford, day n has the property $\sim W$ of failing to be warm and sunny.) In this example, we have stipulated that our body of knowledge, K , includes the premises of both argument (1) and argument (2). Consider argument (1), offered as an explanation of why that November day was warm and sunny. Is argument (1) an admissible I-S explanation of its conclusion, once we add the RMS? The conjunction of the premises of (1) and K implies that n , November 27 in Stanford, belongs to the class $(N \& S)$, which is a subclass of N . Now, either the conjunction of the premises of (1) and K implies the probability

of W (a warm and sunny day) in $(N \& S)$, or it does not. If the value of $P(W/N \& S)$ is not given, then the RMS is not satisfied. In that case, neither (1) nor (2) will qualify as an adequate I-S explanation. Alternatively, if the conjunction of the premises of (1) and K does imply that $P(W/N \& S) = r_1$, then we have to consider what the value of r_1 is. On the one hand, if $r_1 = r$, where r is the numerical value of $P(W/N)$, then S is statistically (and explanatorily) irrelevant to W , and argument (1) satisfies the RMS. If there is no other reference class more specific than N to which n is known to belong, then argument (1) stands as an acceptable I-S explanation of its explanandum, Wn . On the other hand, if $r_1 \neq r$, and if the statement that $P(W/N \& S) = r_1$ is not simply a theorem of mathematical probability theory, then the RMS is not satisfied, and argument (1) is disqualified as an acceptable I-S explanation. In this case, some other argument referring n to the more specific class $(N \& S)$ may constitute an acceptable I-S explanation of Wn only if it, in turn, satisfies the RMS. The reason for the "unless" clause in the RMS is thus easy to appreciate. The class $(N \& W)$ is clearly a subset of N , but $P(W/N \& W) = 1$ by the probability calculus alone, and so the probability of W in the subclass $(N \& W)$ must differ from the probability of W in the class N . Obviously, without the "unless" clause, no inductive argument could ever satisfy the RMS.

As we have seen, Hempel explicitly relativizes the RMS to a particular knowledge situation, K . What is K ? It is the class of all the sentences that are accepted as true by empirical science at a given time. Thus, K could (and quite likely does) contain some false sentences, and the contents of K will change over time.¹² This feature of Hempel's RMS has profound consequences for his concept of an I-S explanation. It means that, for Hempel, there is no such thing as an objective, "correct" inductive explanation independent of the scientific context. By relativizing the RMS to the beliefs of scientists at a given time, Hempel is admitting that inductive explanations (unlike their D-N counterparts) are fundamentally relative and subjective: they depend on the beliefs of scientists for their very existence. Hempel calls this feature of inductive explanations the *epistemic relativity of statistical explanation*.

6.4 | Ruben on the Irrelevance Objection to Hempel's Models of Explanation

Many philosophers of science have judged Hempel's models of explanation unacceptable in one or another respect and have proposed alternative accounts to meet its difficulties. In the remainder of this commentary, we discuss two such accounts and the objections from which they arise. The first, from David-Hillel Ruben, emerges from an important criticism of Hempel's covering law model offered in various forms (by Peter Achinstein, Wesley Salmon, and others) and is known as the *irrelevance objec-*

tion. (The second account, by Peter Railton, is discussed in the final section of this commentary.) After evaluating the proposal that Hempel's account might be repaired by including a description of the explanandum's cause in the explanans—the *causal condition*—Ruben considers reasons for departing from Hempel's insistence that explanations are arguments involving laws.

In the first part of his article "Arguments, Laws, and Explanation," Ruben gives two sets of counterexamples to Hempel's covering law model of explanation. The first set (using Ardon Lyon and Baruch Brody's examples) concerns the explanation of laws; the second set (using Achinstein and Salmon's examples) deals with the explanation of particular facts. We shall focus on the second set.

THE ARSENIC EATER

Achinstein invites us to consider the ill-fated Jones, who eats at least a pound of arsenic and dies within twenty-four hours. Suppose that it is a law of nature that anyone who eats that much arsenic will be dead within a day. From this law and the initial condition that Jones ingests more than a pound of arsenic, we can deduce that Jones dies within twenty-four hours of his eating the arsenic. Thus, we have here an argument that satisfies all the conditions of Hempel's D-N model and that seems to be a good explanation of why Jones died. But then we learn that Jones did not die of arsenic poisoning but was run over by a bus shortly after his poisonous meal. Clearly, the D-N argument citing the lethal properties of arsenic now seems to fail as an explanation. Even though the premises of the D-N argument are true and make essential use of a (true) law to validly entail that Jones dies within twenty-four hours, those premises do not explain why Jones died. Why? Because it was the bus that killed Jones, not the arsenic. The premises of the D-N argument are explanatorily irrelevant to the explanandum. (On pp. 722–24, Ruben considers a number of ways in which the D-N model might, using resources already present in Hempel's theory, be defended against the arsenic-and-bus counterexample and argues that they are unsuccessful.)

THE BIRTH-CONTROL PILLS

It is poor Jones yet again who figures in a counterexample to Hempel's model, this time from Wesley Salmon. John Jones takes birth-control pills regularly and fails to become pregnant. This is hardly surprising. Yet it is, presumably, a law that any man who takes birth-control pills regularly will fail to become pregnant, and John Jones is such a man. So we have a D-N argument with Jones's failure to become pregnant as its conclusion. But clearly this argument does not explain why Jones failed to become pregnant. Jones failed to become pregnant because he is a man, not be-

cause he is a man who took birth-control pills regularly. The explanans of the D-N argument does assert that Jones is a man, but it also includes the irrelevant information about the oral contraceptives. As Ruben argues, it is the presence of this additional, irrelevant information that robs the D-N argument of its explanatory power.

THE HEXED SALT

Salmon and Achinstein's examples can also be adapted to provide counterexamples to Hempel's I-S model. Suppose, for example, that ordinary table salt has a high probability (say, 0.95) of dissolving when stirred into cold water for five minutes. I take some salt and place a "dissolving spell" on it. It is now a sample of hexed salt.¹¹ It is a law that all hexed salt dissolves in water with a probability of 0.95. But although this law can be used to predict that my sample of hexed salt will dissolve in water, it does not explain why it does so. As with Salmon's birth-control pills example, the I-S argument based on the hexed-salt law contains irrelevant information.

Notice that in the hexed-salt example, Hempel's RMS is satisfied. There is no more specific reference class to which the hexed salt could be assigned that would make any difference to its chances of dissolving. Indeed, what has gone wrong in this example is that the salt has already been assigned to a reference class (the class of things that are hexed salt) that is *too* specific. The solution to the problem might seem to lie in a simple modification of Hempel's RMS. Instead of requiring that the explanandum be referred to the most specific class that makes a difference to the probability, we should instead assign the explanandum to the widest, least specific class that satisfies the RMS. Salmon has named this new requirement, *the requirement of the maximal class of maximal specificity*. Unfortunately, this proposal (which Hempel advocated in 1968) will not work.¹⁴

To see why it fails, suppose, for the sake of argument, that sodium bicarbonate has a probability of 0.95 of dissolving when stirred into cold water for five minutes. This is exactly the same as the probability of salt (sodium chloride) dissolving under the same circumstances. Some white powder (that we know to be salt) is stirred into cold water and after five minutes all of it has dissolved. What explains the fact that the powder dissolved? According to the requirement of the maximal class of maximal specificity, we must seek the widest class that satisfies the RMS. The class of things that are salt satisfies the RMS; so does the class of things that are sodium bicarbonate; and so, too, does the disjunctive class of things that are either salt or sodium bicarbonate. On the assumption that we know of no other chemicals that dissolve in water with a probability of 0.95, it follows that the class of things that are either salt or sodium bicarbonate is the widest class that satisfies the RMS. Thus, according to

the requirement of the maximal class of maximal specificity, what explains the dissolving of the powder is not the fact that it is salt but the fact that it is either salt or sodium bicarbonate. On this proposal, to say that the powder is salt is irrelevant—just as irrelevant as saying that it is hexed or that it was mined in Utah. And that just seems wrong. Surely, what explains the dissolving of the powder—at least at some level of explanation—is the fact that it is salt. Thus, while the RMS fails because it permits the explanandum to be referred to classes that are too specific, the new proposal fails because, in some cases, it requires that the explanandum be referred to classes that are too wide.

A PROPOSED CURE FOR THE IRRELEVANCE PROBLEM: THE CAUSAL CONDITION

Reflection on the irrelevance problem (and on the symmetry problem, discussed earlier) leads naturally to the idea that an adequate explanation for a particular fact must include a description of the cause of that fact. As Ruben quotes Salmon as saying, such a proposal would “put the *cause* back into *because*” (728). Thus, we need to consider whether Hempel’s model can be repaired by adding an empirical causal condition—a condition requiring that the explanans contain a description of the cause of the explanandum and that this description play an essential role in the derivation of the conclusion of the explanatory argument.

As a remedy for the irrelevance and symmetry problems, the causal condition sounds promising. But as Timothy McCarthy has shown, and as Ruben explains, the causal condition is no guarantee that the resulting argument will be genuinely explanatory. Consider a version of the first of McCarthy’s counterexamples, regarding an attempt to explain why a particular forest caught fire. Let us suppose that the actual cause of the forest fire was a lightning strike. Our D-N argument (substituting for the As, Bs, Cs, and Ds in McCarthy’s formula) runs as follows:

- 1 All metals are conductors
- 2 The forest was struck by lightning, and this screw is metallic.
- 3 Either this screw is not a conductor, or the forest was not struck by lightning, or the forest caught fire

- 4 The forest caught fire.

Admittedly, this example is highly artificial. But the point is that, even though it satisfies Hempel’s requirements and the causal condition, this argument surely fails to explain why the forest caught fire, because it is circular.

The circularity in the forest fire argument can be diagnosed in the following way. Premise (3) is a disjunctive statement containing three un-

related disjuncts. In order for premise (3) to be true, at least one of its disjuncts must be true. From premises (1) and (2) we can deduce that the first two disjuncts are false. So to know that the third premise is true, we must know that the third disjunct is true. But the third disjunct is simply a restatement of the conclusion. Thus, in a fairly obvious sense, the forest fire argument is viciously circular. It was this problem of vicious circularity that Jaegwon Kim attempted to avoid in the way described by Ruben. Kim imposed yet a further condition, namely, that the explanandum (the conclusion) not entail any of the conjuncts in the singular premises when those premises are written in conjunctive normal form. Kim's condition rules out the forest fire argument because the third premise of that argument is already in conjunctive normal form and is entailed by the explanandum. However, as Ruben explains, McCarthy was able to devise yet a further argument that satisfies Hempel's conditions, the causal condition, and Kim's condition but that still fails to explain its conclusion.

6.5 | Ruben's Single-Statement View of Explanation

The purpose of the causal condition, in meeting the irrelevance and symmetry objections, is to tighten the connection between explanans and explanandum so that their deductive relation exists by virtue of some actual causal connection. But notice that the proposed causal condition merely requires that the premises of an explanatory argument mention the event, *c*, which, as a matter of fact, is the cause of the explanandum event, *e*. The causal condition does not require that the premises contain a statement that says "*c* is the cause of *e*." McCarthy's counterexamples show that, in this form, the causal condition is too weak: a mere mention of the cause does not secure the explanatory relevance we seek. But a suitable strengthening of the causal condition is not easily accomplished. Suppose we strengthen the causal condition so that it now does require a statement in the premises that says "*c* is the cause of *e*." The consequences of this adjustment for the theory of explanation are radical. For it now emerges that all explanations of particular events would reduce to a very simple argument that has but a single premise, namely:

c is the cause of *e*.
e.

As Ruben remarks, this argument is so trivial that the thesis of Hempel (Carnap, Mill, Aristotle, and many others) that explanations are arguments is called into question. For it seems that we could just as well say that the explanation of *e* is the single (true) statement that *c* is the cause of *e*.

Moreover, in such a construal of explanation no law is explicitly mentioned.

It is in this way, then—by considering important objections to Hempel's covering law model and the difficulty of avoiding them with a causal condition—that Ruben is led to a view of explanation that departs significantly from Hempel's model. In particular, this new account of explanation denies two of Hempel's key doctrines: that explanations are arguments and that they must explicitly involve laws.

RUBEN ON THE ROLE OF LAWS IN EXPLANATION

Ruben agrees with Hempel, Carnap, and others that if explanations are arguments, then explanations must include laws. (In a D-N argument explaining a single event, the lawlike generalization is essential if the premises are to entail the conclusion.) But, clearly, it does not follow from this that if explanations are not arguments, then they need not include laws, nor does it follow that laws are irrelevant to explanation. Ruben's own position is that every explanation is a single statement, not an argument. Some explanations (especially those in the physical sciences) do include laws, but the role laws play in such explanations is not that of a premise in an argument. Other explanations can be full and complete without containing laws, although relevant laws do still play an important role in such explanations.

In the final two sections of "Arguments, Laws, and Explanation," Ruben discusses how laws can be relevant to explanation on his single-statement view. On Ruben's view, *o*'s being *F* is the full explanation of why *o* is *G* only if it is a law that all *F*s are *G*, without exception or qualification. If it is false that all *F*s are *G*, then *o*'s being *F* does not fully explain why *o* is *G*. Perhaps the real connection between *F*-ness and *G*-ness is more complicated and the relevant law is: $(x)(Fx \ \& \ Kx \ \& \ Hx \ \& \ Jx \supset Gx)$. The point is not that the explanation of why *o* has property *G* must include the law $(x)(Fx \ \& \ Kx \ \& \ Hx \ \& \ Jx \supset Gx)$. Rather, it is that a full explanation of why *o* is *G* must mention not only that *o* is *F*, but also that *o* is *K*, *H*, and *J* as well. In this way, laws can be relevant to explanations without appearing in them explicitly.

Let us now suppose that it is a law that all *F*s are *G* and that someone claims that *o*'s being *F* is the full explanation of why *o* is *G*. Ruben argues that such a claim seems correct not because of the law (regarded as the generalization that all things that are *F* are also *G*), but because the law calls to our attention the properties that *o* possesses. From this point of view, it is strictly irrelevant that other *F*-things are also *G*-things; what matters, and the only thing that matters, for explaining why *o* is *G* are the properties of *F*-ness and *G*-ness that this particular thing, *o*, happens to have. The explanatory work is done by the properties *o* possesses, not by the law stating that all *F*s are *G*.

As Ruben remarks, it is hard to see why it should be thought relevant in explaining why this thing has property *G* to be told, via a law statement, about *other* things that also have properties *F* and *G*. No doubt some connection between *F*-ness and *G*-ness must hold, but this is precisely what the explanation claim under consideration already expresses. If *o*'s being *F* is the full explanation of why *o* is *G*, then, because of the implicit generality of the properties *F*-ness and *G*-ness, it follows logically that it is a law that all *F*s are *G*. But the law itself is not part of the explanation.

In the final section of his paper, "Generalizations Get Their Revenge," Ruben argues that laws (and theories) are often indirectly relevant to explanation, especially the deep kinds of explanation one finds in physics, by providing scientists with the vocabulary in which to phrase their explanations. What distinguishes a deep explanation from a shallow one is the explanatory power of the predicates that the explanation statement employs.¹⁵ Almost inevitably in such cases, the explanation statement will explicitly include laws that connect the deep, explanatory properties with the more superficial properties used to describe the explanandum. Thus, on the surface at least, a full explanation in the sciences on Ruben's account may end up looking very similar to an ideal Hempelian D-N explanation, but the philosophical interpretation of what this and other explanations really are is fundamentally different.

6.6 | Railton on What Is Wrong with Hempel's Models of Explanation

In his article "A Deductive-Nomological Model of Probabilistic Explanation," Peter Railton begins by discussing what he takes to be the fundamental flaws in Hempel's two models of explanation. After diagnosing these flaws, Railton presents his own account, the D-N-P model of explanation.

NOMIC SUBSUMPTION

Throughout his writings on explanation, Hempel's guiding assumption is that to explain some event or phenomenon is essentially a matter of subsuming the explanandum under a law (whether universal or statistical). For Hempel, explanation just is nomic subsumption. But as we have seen from the earlier discussion of the thesis of structural identity and the symmetry objection criticizing that thesis, not every case of nomic subsumption is an explanation. We can, for example, use the laws of optics and geometry to predict the height of the flagpole from the length of its

shadow and the elevation of the sun, but this prediction does not explain why the flagpole has the height it does. Nomic subsumption is not sufficient for explanation.

According to Railton, what is lacking in D-N arguments that fail to explain is an account of the underlying mechanisms responsible for causing the fact to be explained. Without providing some such account, many D-N arguments are too superficial to explain their conclusions. To explain why this A is also a B, we need to do more than cite the law that all As are Bs. For even if the law in question is causal, the explanation would be incomplete without an account of the mechanism (or mechanisms) at work. This condition, that an explanation provide a mechanism, plays a central role in Railton's effort to give an account of probabilistic explanation. Railton insists that explanations require mechanisms even when the phenomena in question are not deterministic but irreducibly probabilistic.

INDUCTIVE ARGUMENTS

Railton takes issue with two prominent features of Hempel's I-S model of probabilistic explanation: the requirements of high probability and maximal specificity. Railton thinks that both of these objectionable features stem from Hempel's insistence that the statistical explanation of particular facts be a type of *inductive* argument. Railton's proposal, in light of this diagnosis, is simple and radical: probabilistic explanations are not inductive arguments. Indeed, on Railton's D-N-P model, probabilistic explanations are not arguments at all; they include arguments, but none of the arguments they include is inductive, and the explanations are not themselves arguments.

To appreciate Railton's diagnosis of the source of those two features he finds worrisome in Hempel's I-S model, consider, first, the high-probability requirement. An inductive argument's strength depends on the degree of probability that its premises confer on its conclusion—the higher the probability, the stronger the argument. If statistical explanations are inductive arguments, then the high-probability requirement immediately follows: a statistical explanation will be strong only if it establishes that the explanandum occurs with suitably high probability. Now consider the RMS. Inductive arguments are notoriously sensitive to the addition of information to their premises. (In this respect, inductive arguments differ from deductively valid arguments. If an argument is deductively valid, then it remains valid, no matter what is added to its premises.) If Smith is a twenty-year-old woman, and 90 percent of such women survive to celebrate their fortieth birthday, then, relative to those premises, the probability of the conclusion that Smith will live at least another twenty years is quite high. But once we add the information that Smith has AIDS, the probability of her long-term survival plummets. Thus, we generate the problem

of ambiguity that Hempel's RMS is intended to solve. Again, the root of the problem is the inductive nature of the arguments involved.

THE HIGH-PROBABILITY REQUIREMENT

Assuming that Railton has correctly traced the problems with Hempel's high-probability and maximal-specificity requirements to their inductive source, are these features as damaging to Hempel's account as Railton claims? The high-probability requirement has often been criticized (by Salmon, Jeffrey, and others) on the grounds that it rules out the possibility of explaining improbable events. Railton gives the example of a genuinely random wheel of fortune with 99 red stops and 1 black stop. The stipulation that the wheel is genuinely random means that no factor, even in principle, can affect the outcome once the wheel has been set spinning. Each stop thus has exactly the same probability of being chosen when the wheel comes to rest. In a setup like this, it seems absurd to insist that we can explain why the wheel halts at a red stop but not why the wheel halts at the black one. Surely, the explanation is equally good in either case, regardless of the outcome—there is an irreducibly indeterministic mechanism that generates red with probability 99/100 and black with probability 1/100. Hempel's high probability requirement seems to conflate *inductive strength* with *explanatory value*. It is the second that is of concern in seeking an adequate account of explanation. The predictive inference to the conclusion that the wheel will stop at black is much weaker, inductively, than the inference to the conclusion that the wheel will stop at red. But the explanation of why the wheel stops at black is just as good as the explanation of why the wheel stops at red, since, in either case, the explanatory statements provide us with as much of an understanding of the underlying mechanism as it is possible to have. Explanations should be judged by the completeness of the explanatory information they provide, not by the strength of the inferences they permit. To think otherwise is, in Railton's words, to confuse explanation with induction.

THE MAXIMAL SPECIFICITY REQUIREMENT

Railton rejects Hempel's requirement of maximal specificity because it relativizes the concept of probabilistic explanation. If Hempel were right, then there could be no correct or true probabilistic explanation of anything, for all such explanations would be relative to the state of scientific knowledge (and ignorance) at the time they were proposed. For Railton this is unacceptable because it would deprive probabilistic explanation of its objectivity.¹⁶ It is important to note here that Railton is not denying that which explanations we believe to be correct depends on our beliefs. In just the same way, which propositions we believe to be true also depends on our beliefs. But, Railton insists, what it is for something to be a

correct explanation cannot depend on our beliefs and neither can what it is for a proposition to be true. Indeed, it is only if being a correct explanation and being true are belief-independent that we can make sense of our beliefs that particular propositions are true or that particular explanations are correct.

Part of the reason for Hempel's error, Railton thinks, is that Hempel has failed to take seriously his own distinction between statistical descriptions (that just happen to be true) and genuine probabilistic laws. Suppose, on the one hand, that determinism were true. In this case, all laws would be universal; there would be no probabilistic laws and thus no correct or true probabilistic explanations. In such a purely deterministic world, statistical explanations would be merely a stopgap measure until we could discover the true, objective D-N explanations of the things that happen in that world. The statistical "laws" mentioned in such explanations would not be real laws at all but merely expressions of our ignorance at a given time. While it may seem as if, in a completely deterministic world, the conditions for Hempel's I-S model of explanation would be satisfied (capturing our best probabilistic explanations in varying states of ignorance), in fact they would not be satisfied, simply because in such a world there could be no genuine probabilistic laws at all.

On the other hand, suppose that the world is not deterministic but is governed by at least some true probabilistic laws. Once we realize that true probabilistic laws require genuine indeterminism, the epistemic relativity thesis collapses and with it goes the motivation for Hempel's RMS. Remember that the RMS says that, for the purposes of explaining why o has property G , we should assign o to the most specific reference class of which we have knowledge that would make a difference to the probability that o is G . (It is the reference in RMS to our state of knowledge, that is, to our beliefs at a given time, that makes probabilistic explanation epistemically relative on Hempel's account.) But if it is a genuine law that the probability that o is G , given that o is F , is 0.95, then there cannot be any more specific reference class than F to which o could be assigned that would make a lawlike difference to the probability that o is G . If it is a genuine probabilistic law that $P(G/F) = 0.95$, then, if o is F , the real, objective probability that o is G is 0.95, regardless of what anyone thinks or believes. Genuine probabilistic laws require indeterminism, and such indeterminism guarantees the objectivity of statistical explanation.

6.7 | Railton's Deductive-Nomological Model of Probabilistic Explanation

Railton's main concern is the explanation of what he calls *lawful* chance phenomena. Chance phenomena (such as radioactive decay) are lawful when they obey a statistical or probabilistic law. Like Railton, we shall

focus on the explanation of particular events that are brought about by chance mechanisms in a lawful way.

Several key elements of Railton's D-N-P model of explanation have emerged already from the preceding discussion of Railton's critique of Hempel. Those elements can be summarized as follows:

- all explanations are objective; none of them is relative to a set of beliefs or to the state of scientific knowledge at a particular time;
- explanations are not arguments, nor should they be evaluated as if they were arguments, explanations are accounts that provide relevant information;
- explanations (whether probabilistic or not) require not only laws but also an account of the underlying mechanism(s);
- probabilistic explanations require genuine probabilistic laws; genuine probabilistic laws require indeterminism;
- there is no high-probability requirement for probabilistic explanation; improbable events can be explained just as well as highly probable events

RAILTON'S CONDITIONS OF ADEQUACY FOR D-N-P EXPLANATION

Like Hempel's D-N model, Railton's D-N-P model can be set out in schematic form. (We have made minor changes to Railton's notation.)

Explanandum Ge, t_0 : e 's having property G at time t_0

Explanans

a	A theoretical derivation of a probabilistic law of the form (b)	Theoretical derivation
b	$(t)(x) [Fx, t \rightarrow P(Gx, t) = r]$	} Deductive argument
c	Fe, t_0	
d	$P(Ge, t_0) = r$	
e	(Ge, t_0)	Parenthetic addendum

The explanandum, the event to be explained, is e 's having property G at time t_0 , which Railton writes as Ge, t_0 . For example, the explanandum might be that a particular wheel of fortune, e , stopped on black at time t_0 . The statement that this event occurred appears again at the end of the explanans, but only as a parenthetical addendum. Putting Ge, t_0 in parentheses in line (e) is Railton's way of indicating that the explanandum is not the conclusion of the explanans, nor is it inferred from the explanans. Rather, the parenthetical addendum is put there simply to remind us that the explanandum event did, in fact, occur. Remember, although the explanans contains arguments—the deduction of (d) from (b) and (c), plus whatever arguments are involved in the theoretical derivation of the probabilistic law in line (a)—the explanans as a whole is not itself an argument.

The probabilistic law written schematically in line (b) should be read as follows: for all times t , and all things x , if x has property F at time t , then x has probability r of having property G at that time. For example, if F is the property of being a genuinely random wheel of fortune with 99 red stops and 1 black stop, and G is the property of stopping on black after the wheel has been set spinning, then the value of r is $1/100$. (In this example, the probability does not depend on time and so the variable t can be dropped from the statement of the law.)

Line (d) is derived from lines (b) and (c) in two steps: by universal instantiation followed by *modus ponens*. Again, it helps to consider our simple example of the wheel of fortune. The probabilistic law for the wheel of fortune is a universal generalization: it says (ignoring the time variable) that "for any object whatever, if that object is a genuinely random wheel with 99 red stops and 1 black stop, then the probability that the wheel will stop on black is $1/100$." Universal instantiation allows us deduce that this generalization holds for a particular object, such as e . So we get the conditional statement: "if e is a genuinely random wheel with 99 red stops and 1 black stop, then the probability that e will stop on black is $1/100$." In our wheel of fortune example, line (c) tells us that the antecedent of this conditional statement is true: e is indeed a genuinely random wheel with 99 red stops and 1 black stop. Thus, by *modus ponens* we can deduce further that the probability that e will stop on black is $1/100$. As Railton remarks about an exactly similar deduction (argument (2) on page 754), it is vital for the first step of this derivation that (b) be a genuine universal law, for only then can conclusions be deduced from it by universal instantiation. If (b) were merely a statistical generalization saying, in our example, that in a very large sample of N wheels, on average, $N/100$ of them stop on black, nothing could be validly deduced about the probability that this particular wheel e will stop on black.

Line (a) of the explanans expresses Railton's requirement that any

adequate explanation must specify an underlying mechanism that causally brings about the event to be explained. While the derivation of the probabilistic law of the form (b) is indeed meant to issue from scientific theory, Railton has in mind a quite liberal notion of what counts as a mechanism. For example, in the quantum-mechanical derivation of the law of radioactive decay, which Railton gives as his central illustration, the alpha particle escapes from the nucleus by what is called *tunneling* or *barrier penetration*. But tunneling is not a visualizable mechanism or any kind of process that is continuous in space and time. Like other distinctively quantum phenomena, it involves the discontinuous changes in energy and momentum of a system obeying the fundamental laws of quantum mechanics. The alpha particle, for example, only has a distinct existence as a particle, with a well-defined trajectory and momentum, after it has emerged from the nucleus. To refer to tunneling as a mechanism requires using the word *mechanism* in a very broad sense to mean, roughly, *a law-governed process (not necessarily spatially continuous or extended in time) that causally brings about the event we are interested in*. Newton's force of gravity, acting instantaneously at large distances through a vacuum, would qualify as a mechanism. The important thing, for Railton, is the theoretical derivation, not the intuitive idea of a mechanism as an assembly of pulleys, wheels, and strings.

PROPENSITIES AND PROBABILITIES

The deductively valid argument appearing explicitly in a D-N-P explanation is the inference from (b) and (c) to (d). Thus, just as in Hempel's D-N model, this argument contains a statement of law and a statement of initial conditions in its premises. The crucial difference between the two models lies in the nature of the law and the resulting difference in the conclusion. In a D-N-P explanation, the law is probabilistic and expresses the propensity, $P(Gx, t)$, for an object or system, x , to have the property G at time t . What is a propensity? It is a property of the single system x , a lawlike tendency or physical probability for that system to behave in a certain way. The conclusion that $P(Ge, t_0) = r$, which is deduced from (b) and (c), is a single-case probability; it is the probability that this particular e is G at the time in question. This notion of a propensity as a single-case probability is crucial to Railton's model. As Railton says in his paper, "the D-N-P model is viable only if sense can be made of propensities, or of objective, physical, lawful, single-case probabilities by any other name" (761).

We can try to grasp what propensities (or single-case probabilities) are supposed to be by considering the simple example of a perfectly symmetrical coin. Because of the perfectly symmetrical distribution of its mass

and the laws of mechanics, the coin has a physical probability of exactly $\frac{1}{2}$ of yielding a head on any toss. Even if the coin is never tossed, it has that propensity. If the coin is tossed and lands tails, its propensity to yield heads on the next toss remains $\frac{1}{2}$. Propensities are not the same as frequencies; rather, they are properties that explain why we observe certain kinds of outcomes with the frequency that we do. Thus, on Railton's model, if (as we have supposed) the coin has a lawlike, indeterministic tendency of $\frac{1}{2}$ to yield a head on any toss, and if we toss the coin eight times and get, much to our surprise, eight heads in a row, then the full, complete explanation of that outcome is that the coin had a propensity of $(\frac{1}{2})^8$ to behave in that way when tossed. The full explanation of the outcome has been given when the true chance of its occurrence, however low, has been given.

Like probabilities, propensities range in numerical value from 0 to 1. But propensities are not the same as probabilities. The difference between probabilities and propensities can be illustrated in the following example. Suppose that we have a 50-50 mixture of two radioactive elements, A and B, each with a different, known propensity to emit alpha rays. We know how likely it is that the A atoms will emit an alpha ray in a given time interval, and we know how likely it is that the B atoms will emit an alpha ray in the same time interval. Suppose that an alpha ray has been emitted from the mixture. From the information provided we can, using Bayes's theorem, calculate the probability that the alpha ray came from the A atoms, and we can also calculate the probability that the alpha ray came from the B atoms. But neither of these inverse probabilities can be a propensity, for an alpha ray existing at a given time cannot have a propensity to have been produced from a particular source at an earlier time.¹⁷ Propensities are indeterministic causal tendencies. Like causes in general, propensities have a forward temporal direction, from the past and present into the future. Unlike probabilities, propensities can never run from the present to the past. Obviously much more work is needed before we can have a clear notion of what propensities are. Railton's own suggestions are the barest beginnings of this project.

RAILTON'S RESPONSE TO SOME OBJECTIONS

Toward the end of his article, Railton raises several objections to his D-N-P model of explanation and responds to them. One of these objections (the first) is particularly instructive. The criticism is that Railton's model of probabilistic explanation is at once too narrow and too broad. On the one hand, the D-N-P model may be judged too narrow because very few phenomena outside of quantum mechanics are genuinely indeterministic. Sciences such as evolutionary biology, genetics, epidemiology, economics, sociology, fluid mechanics, and meteorology all deal with phe-

nomena that are deterministic but very complicated. Statistics and probability are used by these sciences, but largely because we are ignorant of the many thousands (often, millions) of initial conditions actually determining the behavior in which we are interested. Thus, on Railton's model, none of these uses of probability and statistics would be explanatory; the calculations, predictions, and arguments involved are not D-N-P explanations. On the other hand, Railton's model is accused of being too broad because, if any genuinely indeterministic propensities whatsoever are at work, however minuscule, then a proper D-N-P explanation must take account of them. For example, when an ice cube is placed in a glass of warm water, we expect it to melt. Normally, we would think that we have explained this melting by showing that it follows from the laws of classical thermodynamics (such as the second law stating that, in isolated systems, entropy always increases). But ice and water are composed of molecules, and the molecules of atoms; since atoms are subject to quantum-mechanical laws, there is a small probability that those molecules could move in such a way that the ice does not melt but instead becomes even colder while the water becomes hotter. The probability that this might happen is very, very small, but it is not 0. Thus, the propensity of the ice to melt when placed in water is not exactly 1. (So the second law of thermodynamics is not really a law at all because it is, strictly speaking, false that entropy must always increase.) Consequently, a proper explanation of the melting has to be a D-N-P explanation in which the exact numerical value of the propensity is calculated. (This, of course, is a very difficult task and not one we are normally able to perform.) Moreover, as Railton points out, just about everything that happens in the world, even at the macroscopic level is, to some very small degree, contingent on what happens at the atomic level. So even an explanation of human behavior or the motions of the planets would have to include a fiendishly complicated quantum-mechanical calculation of the relevant propensity before it could qualify as a proper, probabilistic explanation.

Railton's response to both arms of the objection is to concede the point it makes, but to deny that this reflects badly on his model. Thus, Railton willingly embraces the conclusion that if a system is genuinely deterministic, then however complicated it might be, no real explanation of its behavior can be probabilistic. Whatever else we are doing when we appeal to probabilities or statistics in contexts like this—whether it is predicting, approximating, or estimating—it is not explaining, and we should not pretend that it is. Likewise, if the system is indeterministic, to whatever small degree, then we cannot shirk our explanatory responsibility by ignoring that fact. In the final analysis, the aim of explanation is to achieve a true understanding of the ways things really are. Being conscientious about that aim makes the task of explanation—genuine explanation—ex-

tremely difficult. The fault, if fault there be, lies in the complexity of the world, not in the demands of the D-N-P model

6.8 | Summary

In the second half of the twentieth century, most of the philosophical debate concerning the nature of scientific explanation has centered around Hempel's covering law thesis and his two models of explanation, the deductive-nomological (D-N) and the inductive-statistical (I-S). According to the covering law thesis, explanations are arguments (either deductively valid or inductively strong) that have among their premises at least one statement of an empirical law. As Wesley Salmon has pointed out, Hempel's covering law thesis is characteristic of an epistemic conception of scientific explanation, since it takes explanation to be essentially an inference showing that the explanandum event was to be expected. Given the information in the premises of the explanatory argument, the explanandum event could have been predicted, either with certainty (D-N) or with high probability (I-S). This capability of being used as a prediction is seen by many philosophers of science as the hallmark of a good explanation. For Hempel, Carnap, Nagel, and others, being able, at least in principle, to predict the event to be explained guarantees that the explanation has testable, empirical content and gets to the heart of what distinguishes genuine explanation from mere pseudoexplanation.

Hempel's commitment to the epistemic conception of explanation is made explicit in his advocacy of the thesis of structural identity, that all adequate explanations are potentially predictions and that all adequate predictions are potentially explanations. In other words, Hempel insists that there is no formal, objective difference between explanation and prediction. Whether an argument is an explanation or a prediction depends on pragmatic factors, such as when the argument is put forward and the intentions of the scientist who presents the argument. Predictive arguments are advanced prior to the events mentioned in their conclusions either to test the theory from which the laws are taken or, if the theory and its laws are well accepted, to provide a reliable basis upon which to plan for the future. Explanations are advanced after the events mentioned in their conclusions in order to achieve a theoretical understanding of why those events occurred. We make predictions in order to anticipate the future; we give explanations in order to understand the past. But, according to Hempel, in either case, the formal structure of our reasoning is the same and is captured by the D-N and I-S models.

Attacks on the thesis of structural identity (by Michael Scriven and others) go under the general heading of the symmetry objection. They

consist of two kinds of alleged counterexamples: explanations that are not predictions and predictions that are not explanations. Hempel has strenuously opposed the first kind of alleged counterexample, often arguing that the presumed explanation is no such thing. But Hempel concedes that the second subthesis, that all adequate predictions are potentially explanations, is not as secure as its converse.

The irrelevance objection is another important class of criticisms of Hempel's two models of explanation. Various philosophers (including Wesley Salmon and Peter Achinstein) have given arguments that, while they satisfy all of Hempel's conditions for an adequate explanation, do not appear to explain their conclusions. The birth-control pills and hexed-salt examples offer typical cases in which the presence of irrelevant information in the premises robs the arguments of their explanatory power. Because of his commitment to the epistemic conception of explanation, Hempel has denied that these arguments lack explanatory power, but this response seems implausible. What seems to go wrong in these examples is that the lawlike premise in these arguments, while true, fails to identify the cause of the explanandum event. Consequently, Baruch Brody has proposed that we amend Hempel's D-N model by stipulating that the premises of an explanatory argument must contain a description of the event that is the cause of the explanandum. But, as David Hillel-Ruben explains, Brody's proposal has been roundly refuted by counterexamples devised by Timothy McCarthy. As Ruben points out, it is not enough that the premises contain a description of the event that is, as a matter of fact, the cause of the explanandum event; rather, the premises must explicitly identify that event as the cause. But in that case, the underlying structure of explanations is devastatingly simple—it consists simply of the inference from "*c* is the cause of *e*" to the conclusion "*e*"—and explanations need make no explicit mention of any laws. Thus, we are led to the view that Ruben believes is correct, namely, that explanations are not arguments but single statements and that laws, while vitally important for many types of explanation, are not an essential part of the sentence that explains.

Another important class of criticisms of Hempel concern his I-S model of probabilistic explanation. In order to avoid the problem of ambiguity for statistical explanation, Hempel introduced the requirement of maximal specificity (RMS). But the RMS is relativized to what Hempel calls "a given knowledge situation" (715), thus making statistical explanations depend for their very existence on the body of beliefs that scientists accept at a particular time. This epistemic relativity of statistical explanation in the I-S model stands in marked contrast to the objective, nonrelative character of explanations in the D-N model. Moreover, because Hempel construes all explanations as formally equivalent to predictions, I-S arguments are taken to explain their conclusions only when those conclusions follow with high inductive probability from their premises. Since the strength of the inductive relation between premises and conclu-

sion depends directly on the value of the probability that appears in the statistical lawlike premise of such arguments, it follows that, on Hempel's model, it is impossible to explain improbable events. Both of these features—the epistemic relativity of I-S arguments and the high-probability requirement for explanation—have struck critics (such as Salmon and Railton) as serious flaws in Hempel's account.

Peter Railton has proposed his deductive-nomological model of probabilistic explanation (the D-N-P model) as an alternative to Hempel's account. Unlike Hempel's I-S model, Railton's D-N-P model regards probabilistic explanations as fully objective (not relative to any set of scientific beliefs) and permits the explanation of improbable events. Like Ruben, Railton denies that explanations are arguments, although unlike Ruben, Railton thinks that all explanations must contain a deductive argument based on a law. Central to Railton's D-N-P model is the requirement that explanations specify the causal mechanism that brings about the event (or the kind of event) referred to in the explanandum. Genuine probabilistic explanations can be given only when the mechanism at work is indeterministic, and when an indeterministic mechanism is involved, any genuine explanation must be probabilistic. Railton interprets probabilistic laws, not as generalizations about the frequency of certain kinds of event, but as statements of single-case propensities, such as the chance of getting heads when a particular coin is tossed. The propensity of an object or system is its causal tendency to behave in a particular way. Like probabilities, the strength of a propensity can vary on a scale from 0 to 1, but, unlike the empirical frequency interpretation of probability, propensities are just as much physical properties of individual things and systems as their mass and electric charge; in fact Railton sometimes refers to propensities as physical probabilities.

Apart from the difficulty of understanding exactly what propensities are and how they are related to frequencies, Railton's D-N-P model has a number of consequences that may strike one as counterintuitive. Many of these counterintuitive features stem from Railton's demand that explanation properly so-called must be based on the deepest theoretical understanding of nature that we have. Thus, for example, if a process (such as the melting of an ice cube in warm water) involves the slightest chance that the ice cube not melt but become colder, then Railton requires that any explanation of the melting involve a theoretical calculation of the exact propensity (not quite equal to 1) of the cube to melt. This sets a very high standard for probabilistic explanation. Also, as previously noted, on the D-N-P model probabilistic explanations are legitimate only when there is genuine, rock-bottom, physical indeterminacy involved. Thus, any attempt to explain the behavior of complex but ultimately deterministic systems by means of statistical generalizations is ruled out (thereby disqualifying many of the arguments currently accepted as explanatory in the physical, biological, and social sciences).

Because of their emphasis on causation and causal mechanisms (whether deterministic or indeterministic), the models of explanation advocated by Ruben and Railton are good examples of what Salmon has called the *ontic* conception of explanation.¹⁸ Unlike the epistemic approach to explanation, the ontic approach does not regard explanations as arguments (even though they may include or involve arguments). What matters on the ontic account is not whether the explanandum can be predicted with high probability, but whether, regardless of the probability of the explanandum, we can give a correct description of the underlying causal mechanism that brought about the event we wish to explain. The ontic conception of explanation allows us to give complete explanations of events even when those events are highly improbable, as would be the case if nature is governed by indeterministic laws, as quantum mechanics supposes. Seen from the point of view of the ontic conception, the demand for high probability imposed by the epistemic conception ultimately stems from that conception's commitment to determinism as the final truth about the structure of the world. Seen from the point of view of the epistemic conception, Railton's insistence that we calculate objective, true propensities in all cases places an unreasonably high demand on what counts as an explanation. The epistemic approach (at least in Hempel's version of it) permits genuine statistical explanations even if the world is deterministic through and through, but it makes their status as explanations relative to human beliefs at a particular time. The ontic approach (in Railton's version) takes probabilistic explanations to be just as fully objective and nonrelative as those based on deterministic laws, but insists that such explanations can be given only when the mechanisms involved are fundamentally indeterministic.

■ | Notes

1. Carl G. Hempel and Paul Oppenheim, "Studies in the Logic of Explanation," *Philosophy of Science* 15 (1948): 567–79. Reprinted, with a postscript, in *Aspects of Scientific Explanation* (New York: Free Press, 1965), 245–95.
2. Carl G. Hempel, "Deductive-Nomological vs. Statistical Explanation," in *Minnesota Studies in the Philosophy of Science*, vol. 3, ed. H. Feigl and G. Maxwell, (Minneapolis: University of Minnesota Press, 1962), 98–169.
3. "A physical theory is not an explanation. It is a system of mathematical propositions, deduced from a small number of principles, which aim to represent as simply, as completely, and as exactly as possible a set of experimental laws" (Pierre Duhem, *The Aim and Structure of Physical Theory*, trans. Philip P. Wiener [1914; Princeton, N.J.: Princeton University Press, 1954], 19).
4. *Aim and Structure*, p. 7.

5. Before judging too harshly Mach and Duhem's skepticism about atoms and the aether, we should remember that the atoms we now believe in are utterly different from those described in the classical "billiard-ball" models of the nineteenth century and that the aether has been discarded entirely. What remains of the kinetic theory of gases and the wave theory of light of the nineteenth century are testable equations and laws, and these alone, according to the positivists, have permanent scientific value.

6. Strictly speaking, then, (R2) should be phrased in terms of lawlike sentences, rather than laws. It is commonly assumed that we can judge whether a sentence is lawlike simply by examining its logical form and the predicates it uses. But to judge that a lawlike sentence is true—and hence expresses a law—requires empirical information.

7. Since the explanans can include laws of mathematics, some of the sentences it contains might not be empirical. But (R3) guarantees that every explanans must contain at least one empirical sentence.

8. Michael Scriven, "Explanation and Prediction in Evolutionary Theory," *Science* 130 (1959): 477–82.

9. See "Popper and the Theory of Evolution" in the commentary on chapter 1.

10. There is no problem of ambiguity for D-N explanations. If a set of true premises deductively entails that *a* is *G*, then no other set of true premises can deductively entail that *a* is not *G*.

11. Carl G. Hempel, "Inductive Inconsistencies," *Synthese* 12 (1960): 439–69. Reprinted in *Aspects of Scientific Explanation*, 53–79.

12. Hempel is using the term *knowledge* in a common, nontechnical sense to mean, roughly, *justified belief*. It is in this sense that we talk of the current state of scientific knowledge while recognizing that some of the beliefs in that body of knowledge are quite likely to be false. When epistemologists use the term in its technical sense, they insist that truth is a necessary condition for knowledge.

13. The example and the phrase *hexed salt* were first proposed by Henry Kyburg. See Henry E. Kyburg, "Comments," *Philosophy of Science* 32 (1965): 147–51.

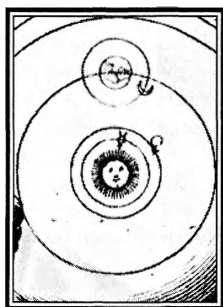
14. Carl G. Hempel, "Maximal Specificity and Lawlikeness in Probabilistic Explanation," *Philosophy of Science* 35 (1968): 116–33. The following criticism of the proposal is adapted from John Meixner, "Homogeneity and Explanatory Depth," *Philosophy of Science* 46 (1979): 366–81.

15. Ruben does not use the phrase "explanatory power." Instead, he talks about laws and theories providing a deeper vocabulary, a vocabulary that "gives a more profound insight into the phenomenon at hand" (741). One way of understanding this notion would be to associate the deeper vocabulary with those theories that unify many different phenomena within a single framework. But since Ruben insists that unification and explanation are two different things, he would reject this suggestion.

16. This line of criticism is also advanced in J. Alberto Coffa, "Hempel's Ambiguity," *Synthese* 28 (1974): 141-63.

17. This criticism originated with Paul Humphreys. See his "Why Propensities Cannot Be Probabilities," *Philosophical Review* 94 (1985): 557-70.

18. Wesley C. Salmon, *Scientific Explanation and the Causal Structure of the World* (Princeton, N.J.: Princeton University Press, 1984), chs. 1 and 4.



7

|

Laws of Nature

INTRODUCTION

Laws play a central role in scientific reasoning. As we saw in chapters 1 and 4, some philosophers of science think that using laws to explain things is an essential part of what it means to be genuinely scientific, and support for the view that scientific explanation must involve laws is widespread (though not unanimous). Many also believe that we are justified in trusting scientific inferences because these predictions rest, in part, on well-confirmed laws. Our expectations about the behavior of systems, instruments, and materials are reasonable to the extent that they are based on a correct understanding of the laws that govern them. Undoubtedly, much scientific activity is devoted to discovering laws, and one of the most cherished forms of scientific immortality is to join the ranks of Boyle, Newton, and Maxwell by having a law (equation or functional relation) linked with one's name. But despite the crucial importance of laws in science, it is difficult to find a general account of what sort of things laws are that can do justice to everything we take to be true of them.

In this chapter, two important and influential ways of understanding laws—the regularity approach and the necessitarian approach—will be discussed and criticized.¹ In terms roughly hewn, the regularity approach says that laws describe the way things actually behave, that they are nothing more than a special kind of descriptive summary of what has happened and what will happen. The necessitarian approach insists that laws are more than just summaries, that they tell us not merely how things actually behave, but, more importantly, how they must behave. For the necessitarians, both the universality and the necessity of laws are objective, real features of the world (although necessitarians disagree among themselves about the nature of that necessity).²

Modern adherents of the regularity approach trace their origins back

to David Hume and his constant conjunction analysis of causation. In "What Is a Law of Nature?" A. J. Ayer gives a sympathetic account of the epistemological considerations that drove Hume to deny that causal necessity is objective and hence to espouse the simple version of the regularity theory of laws. According to this simple version, a law of nature is nothing more, objectively, than a true universal generalization. Ayer explains the severe problems that afflict the simple Humean theory, including the problem of laws that lack instances and the problem of distinguishing between those generalizations that are genuine laws and those that are true merely by accident. Ayer's tentative solution to these problems is to add epistemic conditions to the regularity analysis of lawlikeness. Thus, according to Ayer's epistemic regularity theory, a law is a true universal generalization about which we have certain beliefs and attitudes and that plays a characteristic role in science.

In "Laws of Nature," Fred Dretske deals what he considers a lethal blow to Ayer's epistemic regularity theory. In its place, Dretske advocates a theory according to which laws are (express or describe) relations of necessitation between universals. Thus, instead of regarding laws as generalizations about events, Dretske regards them as singular statements about the properties events share. Dretske shows how his universals theory of laws can solve several of the difficulties facing the regularity theory.

Like other recent advocates of the universals theory of laws, and unlike earlier proponents of the necessitarian approach, Dretske insists that laws of nature are contingent, not necessary. This creates difficulties for Dretske's theory, since it requires that the nomic necessitation relation between universals hold contingently, not necessarily, and it is hard to see how merely contingent relations could obtain among abstract entities such as universals. One possible response to this problem is suggested by Saul Kripke and Hilary Putnam, who use their new theory of reference to argue that many laws of nature are not contingent but metaphysically necessary. D. H. Mellor criticizes the Kripke-Putnam argument in "Necessities and Universals in Natural Laws." Mellor also attacks the universals theory on the grounds that it cannot accommodate laws that have no instances. In this way, Mellor attempts to cast the regularity theory in a more favorable light by revealing the deficiencies of its rivals.

Despite their disagreement about whether laws involve an element of necessity, the regularity and necessitarian approaches share the conviction that laws of nature describe important facts about reality. That realist assumption about laws is challenged by Nancy Cartwright in the final piece in this chapter. In "Do the Laws of Physics State the Facts?" Cartwright argues that most of the laws physicists use to explain things are not even approximately true. They are false and are known to be false. Nonetheless, they provide excellent explanations. Cartwright argues that there is an irreconcilable tension between the goal of accurate description and the

goal of explanation. When lawlike statements are altered to make them describe the way bodies actually behave, they lose their explanatory power.

■ | Notes

1. An older approach to understanding laws of nature—instrumentalism—has largely fallen into disfavor (though Ronald Giere and Bas van Fraassen have recently made attempts to revive it). Instrumentalists (such as Ernst Mach, Karl Pearson, Ludwig Wittgenstein, and Stephen Toulmin) hold that laws are neither true nor false; they are simply tools that scientists use to summarize data and to make inferences. Gilbert Ryle once described this view by characterizing laws as “inference tickets.” According to instrumentalists, neither the necessity nor the universality of laws are objective features of the world; both are human inventions that we impose on the world for the purposes of representation and prediction. The main problem with instrumentalism is that, if laws are neither true nor false, then it is difficult to make sense of their being tested, confirmed, and refuted. See Ronald N. Giere, “The Skeptical Perspective: Science without Laws of Nature,” in *Laws of Nature: Essays on the Philosophical, Scientific and Historical Dimensions*, ed. Friedel Weinert, (New York: Walter de Gruyter, 1995), 120–38; Bas C. van Fraassen, *Laws and Symmetry* (Oxford: Clarendon Press, 1989); and Alan Musgrave, “Wittgensteinian Instrumentalism,” *Theoria* 47 (1981): 65–105.

2. Regularity theorists (of different types) include A. J. Ayer, R. B. Braithwaite, Rudolf Carnap, Richard Feynman, Carl Hempel, Ernest Nagel, Hans Reichenbach, Norman Swartz, and Peter Urbach. Necessitarians (of different types) include D. M. Armstrong, John Bigelow, John Carroll, Fred Dretske, W. C. Kneale, Christopher Swoyer, and Michael Tooley. See the bibliography at the end of this volume for references.

What Is a Law of Nature?

There is a sense in which we know well enough what is ordinarily meant by a law of nature. We can give examples. Thus it is, or is believed to be, a law of nature that the orbit of a planet around the sun is an ellipse, or that arsenic is poisonous, or that the intensity of a sensation is proportionate to the logarithm of the stimulus, or that there are 303,000,000,000,000,000,000,000,000 molecules in one gram of hydrogen. It is not a law of nature, though it is necessarily true, that the sum of the angles of a Euclidean triangle is 180 degrees, or that all the presidents of the third French Republic were male, though this is a legal fact in its way, or that all the cigarettes which I now have in my cigarette case are made of Virginian tobacco, though this again is true and, given my tastes, not wholly accidental. But while there are many such cases in which we find no difficulty in telling whether some proposition, which we take to be true, is or is not a law of nature, there are cases where we may be in doubt. For instance, I suppose that most people take the laws of nature to include the first law of thermodynamics, the proposition that in any closed physical system the sum of energy is constant: but there are those who maintain that this principle is a convention, that it is interpreted in such a way that there is no logical possibility of its being falsified, and for this reason they may deny that it is a law of nature at all. There are two questions at issue in a case of this sort: first, whether the principle under discussion is in fact a convention, and secondly whether its being a convention, if it is one, would disqualify it from being a law of nature. In the same way, there may be a dispute whether statistical generalizations are to count as laws of nature, as distinct from the dispute whether certain generalizations, which have been taken to be laws of nature, are in fact

FROM *The Concept of a Person* (New York: St. Martin's Press, 1963), 209–34. This article was first published in *Revue Internationale de Philosophie* 36 (1956), 144–65.

statistical. And even if we were always able to tell, in the case of any given proposition, whether or not it had the form of a law of nature, there would still remain the problem of making clear what this implied.

The use of the word 'law', as it occurs in the expression 'laws of nature', is now fairly sharply differentiated from its use in legal and moral contexts: we do not conceive of the laws of nature as imperatives. But this was not always so. For instance, Hobbes in his *Leviathan* lists fifteen 'laws of nature' of which two of the most important are that men 'seek peace, and follow it' and 'that men perform their covenants made': but he does not think that these laws are necessarily respected. On the contrary, he holds that the state of nature is a state of war, and that covenants will not in fact be kept unless there is some power to enforce them. His laws of nature are like civil laws except that they are not the commands of any civil authority. In one place he speaks of them as 'dictates of Reason' and adds that men improperly call them by the name of laws: 'for they are but conclusions or theorems concerning what conduceth to the conservation and defence of themselves: whereas Law, properly, is the word of him, that by right hath command over others'. 'But yet,' he continues, 'if you consider the same Theorems, as delivered in the word of God, that by right commandeth all things; then they are properly called Laws.'

It might be thought that this usage of Hobbes was so far removed from our own that there was little point in mentioning it, except as a historical curiosity; but I believe that the difference is smaller than it appears to be. I think that our present use of the expression 'laws of nature' carries traces of the conception of Nature as subject to command. Whether these commands are conceived to be those of a personal deity or, as by the Greeks, of an impersonal fate, makes no difference here. The point, in either case, is that the sovereign is thought to be so powerful that its dictates are bound to be obeyed. It is not as in Hobbes's usage a question of moral duty or of prudence, where the subject has freedom to err. On the view which I am now considering, the commands which are issued to Nature are delivered with such authority that it is impossible that she should disobey them. I do not claim that this view is still prevalent; at least not that it is explicitly held. But it may well have contributed to the persistence of the feeling that there is some form of necessity attaching to the laws of nature, a necessity which, as we shall see, it is extremely difficult to pin down.

In case anyone is still inclined to think that the laws of nature can be identified with the commands of a superior being, it is worth pointing out that this analysis cannot be correct. It is already an objection to it that it burdens our science with all the uncertainty of our metaphysics, or our theology. If it should turn out that we had no good reason to believe in the existence of such a superior being, or no good reason to believe that he issued any commands, it would follow, on this analysis, that we should not be entitled to believe that there were any laws of nature. But the main

argument against this view is independent of any doubt that one may have about the existence of a superior being. Even if we knew that such a one existed, and that he regulated nature, we still could not identify the laws of nature with his commands. For it is only by discovering what were the laws of nature that we could know what form these commands had taken. But this implies that we have some independent criteria for deciding what the laws of nature are. The assumption that they are imposed by a superior being is therefore idle, in the same way as the assumption of providence is idle. It is only if there are independent means of finding out what is going to happen that one is able to say what providence has in store. The same objection applies to the rather more fashionable view that moral laws are the commands of a superior being; but this does not concern us here.

There is, in any case, something strange about the notion of a command which it is impossible to disobey. We may be sure that some command will never in fact be disobeyed. But what is meant by saying that it cannot be? That the sanctions which sustain it are too strong? But might not one be so rash or so foolish as to defy them? I am inclined to say that it is in the nature of commands that it should be possible to disobey them. The necessity which is ascribed to these supposedly irresistible commands belongs in fact to something different: it belongs to the laws of logic. Not that the laws of logic cannot be disregarded; one can make mistakes in deductive reasoning, as in anything else. There is, however, a sense in which it is impossible for anything that happens to contravene the laws of logic. The restriction lies not upon the events themselves but on our method of describing them. If we break the rules according to which our method of description functions, we are not using it to describe anything. This might suggest that the events themselves really were disobeying the laws of logic, only we could not say so. But this would be an error. What is describable as an event obeys the laws of logic: and what is not describable as an event is not an event at all. The chains which logic puts upon nature are purely formal: being formal they weigh nothing, but for the same reason they are indissoluble.

From thinking of the laws of nature as the commands of a superior being, it is therefore only a short step to crediting them with the necessity that belongs to the laws of logic. And this is in fact a view which many philosophers have held. They have taken it for granted that a proposition could express a law of nature only if it stated that events, or properties, of certain kinds were necessarily connected; and they have interpreted this necessary connection as being identical with, or closely analogous to, the necessity with which the conclusion follows from the premisses of a deductive argument; as being, in short, a logical relation. And this has enabled them to reach the strange conclusion that the laws of nature can, at least in principle, be established independently of experience: for if they are purely logical truths, they must be discoverable by reason alone.

The refutation of this view is very simple. It was decisively set out by Hume. 'To convince us', he says, 'that all the laws of nature and all the operations of bodies, without exception, are known only by experience, the following reflections may, perhaps, suffice. Were any object presented to us, and were we required to pronounce concerning the effect, which will result from it, without consulting past observation: after what manner, I beseech you, must the mind proceed in this operation? It must invent or imagine some event, which it ascribes to the object as its effect: and it is plain that this invention must be entirely arbitrary. The mind can never find the effect in the supposed cause, by the most accurate scrutiny and examination. For the effect is totally different from the cause, and consequently can never be discovered in it.'²

Hume's argument is, indeed, so simple that its purport has often been misunderstood. He is represented as maintaining that the inference of an effect in its cause is something which is not discoverable in nature; that as a matter of fact our observations fail to reveal the existence of any such relation: which would allow for the possibility that our observations might be at fault. But the point of Hume's argument is not that the relation of necessary connection which is supposed to conjoin distinct events is not in fact observable: it is that there could not be any such relation, not as a matter of fact but as a matter of logic. What Hume is pointing out is that if two events are distinct, they are distinct: from a statement which does no more than assert the existence of one of them it is impossible to deduce anything concerning the existence of the other. This is, indeed, a plain tautology. Its importance lies in the fact that Hume's opponents denied it. They wished to maintain both that the events which were coupled by the laws of nature were logically distinct from one another, and that they were united by a logical relation. But this is a manifest contradiction. Philosophers who hold this view are apt to express it in a form which leaves the contradiction latent: it was Hume's achievement to have brought it clearly to light.

In certain passages Hume makes his point by saying that the contradictory of any law of nature is at least conceivable; he intends thereby to show that the truth of the statement which expresses such a law is an empirical matter of fact and not an *a priori* certainty. But to this it has been objected that the fact that the contradictory of a proposition is conceivable is not a decisive proof that the proposition is not necessary. It may happen, in doing logic or pure mathematics, that one formulates a statement which one is unable either to prove or disprove. Surely in that case both the alternatives of its truth and falsehood are conceivable. Professor W. C. Kneale, who relies on this objection,³ cites the example of Goldbach's conjecture that every even number greater than two is the sum of two primes. Though this conjecture has been confirmed so far as it has been tested, no one yet knows for certain whether it is true or false: no proof has been discovered either way. All the same, if it is true, it is

necessarily true, and if it is false, it is necessarily false. Suppose that it should turn out to be false. We surely should not be prepared to say that what Goldbach had conjectured to be true was actually inconceivable. Yet we should have found it to be the contradictory of a necessary proposition. If we insist that this does prove it to be inconceivable, we find ourselves in the strange position of having to hold that one of two alternatives is inconceivable, without our knowing which.

I think that Professor Kneale makes his case: but I do not think that it is an answer to Hume. For Hume is not primarily concerned with showing that a given set of propositions, which have been taken to be necessary, are not so really. This is only a possible consequence of his fundamental point that 'there is no object which implies the existence of any other if we consider these objects in themselves, and never look beyond the idea which we form of them',¹ in short, that to say that events are distinct is incompatible with saying that they are logically related. And against this Professor Kneale's objection has no force at all. The most that it could prove is that, in the case of the particular examples that he gives, Hume might be mistaken in supposing that the events in question really were distinct: in spite of the appearances to the contrary, an expression which he interpreted as referring to only one of them might really be used in such a way that it included a reference to the other.

But is it not possible that Hume was always so mistaken; that the events, or properties, which are coupled by the laws of nature never are distinct? This question is complicated by the fact that once a generalization is accepted as a law of nature it tends to change its status. The meanings which we attach to our expressions are not completely constant: if we are firmly convinced that every object of a kind which is designated by a certain term has some property which the term does not originally cover, we tend to include the property in the designation; we extend the definition of the object, with or without altering the words which refer to it. Thus, it was an empirical discovery that loadstones attract iron and steel for someone who uses the word 'loadstone' only to refer to an object which has a certain physical appearance and constitution, the fact that it behaves in this way is not formally deducible. But, as the word is now generally used, the proposition that loadstones attract iron and steel is analytically true: an object which did not do this would not properly be called a loadstone. In the same way, it may have become a necessary truth that water has the chemical composition H_2O . But what then of heavy water which has the composition D_2O ? Is it not really water? Clearly this question is quite trivial. If it suits us to regard heavy water as a species of water, then we must not make it necessary that water consists of H_2O . Otherwise, we may. We are free to settle the matter whichever way we please.

Not all questions of this sort are so trivial as this. What, for example, is the status in Newtonian physics of the principle that the acceleration of a body is equal to the force which is acting on it divided by its mass?

If we go by the text-books in which 'force' is defined as the product of mass and acceleration, we shall conclude that the principle is evidently analytic. But are there not other ways of defining force which allow this principle to be empirical? In fact there are, but as Henri Poincaré has shown,⁵ we may then find ourselves obliged to treat some other Newtonian principle as a convention.* It would appear that in a system of this kind there is likely to be a conventional element, but that, within limits, we can situate it where we choose. What is put to the test of experience is the system as a whole.

This is to concede that some of the propositions which pass for laws of nature are logically necessary, while implying that it is not true of all of them. But one might go much further. It is at any rate conceivable that at a certain stage the science of physics should become so unified that it could be wholly axiomatized: it would attain the status of a geometry in which all the generalizations were regarded as necessarily true. It is harder to envisage any such development in the science of biology, let alone the social sciences, but it is not theoretically impossible that it should come about there too. It would be characteristic of such systems that no experience could falsify them, but their security might be sterile. What would take the place of their being falsified would be the discovery that they had no empirical application.

The important point to notice is that, whatever may be the practical or aesthetic advantages of turning scientific laws into logically necessary truths, it does not advance our knowledge, or in any way add to the security of our beliefs. For what we gain in one way, we lose in another. If we make it a matter of definition that there are just so many million molecules in every gram of hydrogen, then we can indeed be certain that every gram of hydrogen will contain that number of molecules: but we must become correspondingly more doubtful, in any given case, whether what we take to be a gram of hydrogen really is so. The more we put into our definitions, the more uncertain it becomes whether anything satisfies them: this is the price that we pay for diminishing the risk of our laws being falsified. And

* See chapter 6 of *La science et l'hypothèse* (Paris: E. Flammarion, 1902); *Science and Hypothesis*, trans. W. J. Greenstreet (New York: Dover, 1952). Poincaré reasons that any attempt to verify the second law, $F = ma$, by experiment—even on a single body of constant mass—requires a way of measuring forces independently of the accelerations they cause and of ascertaining when two forces are equal in magnitude. This, he argues, must presuppose the truth of the third law (that action and reaction are equal and opposite). Thus, he concludes that if the second law is empirical, then the third law must be treated as a definition. Poincaré also argues that if the second law is treated not as an empirical law but as a definition of force, then it can be applied to more than one body only if the masses of different bodies can be compared. This, too, he argues, presupposes Newton's third law, since when two bodies act on each other, the ratio of their masses is defined as the inverse ratio of their accelerations (assuming that no other bodies are acting on them).

if it ever came to the point where all the 'laws' were made completely secure by being treated as logically necessary, the whole weight of doubt would fall upon the statement that our system had application. Having deprived ourselves of the power of expressing empirical generalizations, we should have to make our existential statements do the work instead.

If such a stage were reached, I am inclined to say that we should no longer have a use for the expression 'laws of nature', as it is now understood. In a sense, the tenure of such laws would still be asserted: they would be smuggled into the existential propositions. But there would be nothing in the system that would count as a law of nature: for I take it to be characteristic of a law of nature that the proposition which expresses it is not logically true. In this respect, however, our usage is not entirely clear-cut. In a case where a sentence has originally expressed an empirical generalization, which we reckon to be a law of nature, we are inclined to say that it still expresses a law of nature, even when its meaning has been so modified that it has come to express an analytic truth. And we are encouraged in this by the fact that it is often very difficult to tell whether this modification has taken place or not. Also, in the case where some of the propositions in a scientific system play the rôle of definitions, but we have some freedom in deciding which they are to be, we tend to apply the expression 'laws of nature' to any of the constituent propositions of the system, whether or not they are analytically true. But here it is essential that the system as a whole should be empirical. If we allow the analytic propositions to count as laws of nature, it is because they are carried by the rest.

Thus to object to Hume that he may be wrong in assuming that the events between which his causal relations hold are 'distinct existences' is merely to make the point that it is possible for a science to develop in such a way that axiomatic systems take the place of natural laws. But this was not true of the propositions with which Hume was concerned, nor is it true, in the main, of the sciences of to-day. And in any case Hume is right in saying that we cannot have the best of both worlds; if we want our generalizations to have empirical content, they cannot be logically secure; if we make them logically secure, we rob them of their empirical content. The relations which hold between things, or events, or properties, cannot be both factual and logical. Hume himself spoke only of causal relations, but his argument applies to any of the relations that science establishes, indeed to any relations whatsoever.

It should perhaps be remarked that those philosophers who still wish to hold that the laws of nature are 'principles of necessitation'⁶ would not agree that this came down to saying that the propositions which expressed them were analytic. They would maintain that we are dealing here with relations of objective necessity, which are not to be identified with logical entailments, though the two are in certain respects akin. But what are these relations of objective necessity supposed to be? No explanation is

given except that they are just the relations that hold between events, or properties, when they are connected by some natural law. But this is simply to restate the problem; not even to attempt to solve it. It is not as if this talk of objective necessity enabled us to detect any laws of nature. On the contrary it is only *ex post facto*, when the existence of some connection has been empirically tested, that philosophers claim to see that it has this mysterious property of being necessary. And very often what they do 'see' to be necessary is shown by further observation to be false. This does not itself prove that the events which are brought together by a law of nature do not stand in some unique relation. If all attempts at its analysis fail, we may be reduced to saying that it is *sui generis* [altogether unique]. But why then describe it in a way which leads to its confusion with the relation of logical necessity?

A further attempt to link natural with logical necessity is to be found in the suggestion that two events E and I are to be regarded as necessarily connected when there is some well-established universal statement U, from which, in conjunction with the proposition *i*, affirming the existence of I, a proposition *e*, affirming the existence of E, is formally deducible.⁷ This suggestion has the merit of bringing out the fact that any necessity that there may be in the connection of two distinct events comes only through a law. The proposition which describes 'the initial conditions' does not by itself entail the proposition which describes the 'effect': it does so only when it is combined with a causal law. But this does not allow us to say that the law itself is necessary. We can give a similar meaning to saying that the law is necessary by stipulating that it follows, either directly or with the help of certain further premisses, from some more general principle. But then what is the status of these more general principles? The question what constitutes a law of nature remains, on this view, without an answer.

■ | II

Once we are rid of the confusion between logical and factual relations, what seems the obvious course is to hold that a proposition expresses a law of nature when it states what invariably happens. Thus, to say that unsupported bodies fall, assuming this to be a law of nature, is to say that there is not, never has been, and never will be a body that being unsupported does not fall. The 'necessity' of a law consists, on this view, simply in the fact that there are no exceptions to it.

It will be seen that this interpretation can also be extended to statistical laws. For they too may be represented as stating the existence of certain constancies in nature: only, in their case, what is held to be constant is the proportion of instances in which one property is conjoined with an-

other or, to put it in a different way, the proportion of the members of one class that are also members of another. Thus it is a statistical law that when there are two genes determining a hereditary property, say the colour of a certain type of flower, the proportion of individuals in the second generation that display the dominant attribute, say the colour white as opposed to the colour red, is three quarters. There is, however, the difficulty that one does not expect the proportion to be maintained in every sample. As Professor R. B. Braithwaite has pointed out, 'when we say that the proportion (in a non-literal sense) of the male births among births is 51 per cent, we are not saying of any particular class of births that 51 per cent are births of males, for the actual proportion might differ very widely from 51 per cent in a particular class of births, or in a number of particular classes of births, without our wishing to reject the proposition that the proportion (in the nonliteral sense) is 51 per cent.'⁸ All the same the 'non-literal' use of the word 'proportion' is very close to the literal use. If the law holds, the proportion must remain in the neighbourhood of 51 per cent, for any sufficiently large class of cases: and the deviations from it which are found in selected sub-classes must be such as the application of the calculus of probability would lead one to expect. Admittedly, the question what constitutes a sufficiently large class of cases is hard to answer. It would seem that the class must be finite, but the choice of any particular finite number for it would seem also to be arbitrary. I shall not, however, attempt to pursue this question here. The only point that I here wish to make is that a statistical law is no less 'lawlike' than a causal law. Indeed, if the propositions which express causal laws are simply statements of what invariably happens, they can themselves be taken as expressing statistical laws, with ratios of 100 per cent. Since a 100 per cent ratio, if it really holds, must hold in every sample, these 'limiting cases' of statistical laws escape the difficulty which we have just remarked on. If henceforth we confine our attention to them, it is because the analysis of 'normal' statistical laws brings in complications which are foreign to our purpose. They do not affect the question of what makes a proposition lawlike: and it is in this that we are mainly interested.

On the view which we have now to consider, all that is required for there to be laws in nature is the existence of *de facto* constancies. In the most straightforward case, the constancy consists in the fact that events, or properties, or processes of different types are invariably conjoined with one another. The attraction of this view lies in its simplicity: but it may be too simple. There are objections to it which are not easily met.

In the first place, we have to avoid saddling ourselves with vacuous laws. If we interpret statements of the form 'All S is P' as being equivalent, in Russell's notation, to general implications of the form ' $(x)(\Phi x \supset \Psi x)$ ', we face the difficulty that such implications are considered to be true in

all cases in which their antecedent is false.* Thus we shall have to take it as a universal truth both that all winged horses are spirited and that all winged horses are tame; for assuming, as I think we may, that there never have been or will be any winged horses, it is true both that there never have been or will be any that are not spirited, and that there never have been or will be any that are not tame.† And the same will hold for any other property that we care to choose. But surely we do not wish to regard the ascription of any property whatsoever to winged horses as the expression of a law of nature.

The obvious way out of this difficulty is to stipulate that the class to which we are referring should not be empty. If statements of the form 'All S is P' are used to express laws of nature, they must be construed as entailing that there are S's. They are to be treated as the equivalent, in Russell's notation, of the conjunction of the propositions ' $(x)(\Phi x \supset \Psi x)$ ' and ' $(\exists x)\Phi x$ '. But this condition may be too strong. For there are certain cases in which we do wish to take general implications as expressing laws of nature, even though their antecedents are not satisfied. Consider, for example, the Newtonian law that a body on which no forces are acting continues at rest or in uniform motion along a straight line. It might be argued that this proposition was vacuously true, on the ground that there are in fact no bodies on which no forces are acting; but it is not for this reason that it is taken as expressing a law. It is not interpreted as being vacuous. But how then does it fit into the scheme? How can it be held to be descriptive of what actually happens?

What we want to say is that if there *were* any bodies on which no forces were acting then they *would* behave in the way that Newton's law prescribes. But we have not made any provision for such hypothetical cases: according to the view which we are now examining, statements of law cover only what is actual, not what is merely possible. There is, however, a way in which we can still fit in such 'non-stantial' laws. As Professor C. D. Broad has suggested,⁹ we can treat them as referring not to hypothetical objects, or events, but only to the hypothetical consequences of instancial laws. Our Newtonian law can then be construed as implying

* Throughout this reading, we have added parentheses to Ayer's formulas. The universal generalization " $(x)(\Phi x \supset \Psi x)$ " should be read as "for all x , if x has property Φ , then x has property Ψ ." Because of the way that the truth-functional connective " \supset " is defined, any conditional formula of the form " $(p \supset q)$ " is true whenever its antecedent, p , is false, regardless of whether the consequent, q , is true or false. Hence, Ayer's remark about winged horses in the next sentence.

† In predicate logic, " $(x)(\Phi x \supset \Psi x)$ " is logically equivalent to " $\sim(\exists x)(\Phi x \ \& \ \sim \Psi x)$." This negation of an existential generalization says "it is not the case that there exists anything, x , such that x has property Φ and lacks property Ψ ." Consequently, when nothing has property Φ —as in Ayer's example of winged horses—both statements are true, regardless of the nature of property Ψ .

that there are instancial laws, in this case laws about the behaviour of bodies on which forces are acting, which are such that when combined with the proposition that there are bodies on which no forces are acting, they entail the conclusion that these bodies continue at rest, or in uniform motion along a straight line. The proposition that there are such bodies is false, and so, if it is interpreted existentially, is the conclusion, but that does not matter. As Broad puts it, 'what we are concerned to assert is that this false conclusion is a necessary consequence of the conjunction of a certain false instancial supposition with certain true instancial laws of nature'.

This solution of the present difficulty is commendably ingenious, though I am not sure that it would always be possible to find the instancial laws which it requires. But even if we accept it, our troubles are not over. For, as Broad himself points out, there is one important class of cases in which it does not help us. These cases are those in which one measurable quantity is said to depend upon another, cases like that of the law connecting the volume and temperature of a gas under a given pressure, in which there is a mathematical function which enables one to calculate the numerical value of either quantity from the value of the other. Such laws have the form ' $x = Fy$ ', where the range of the variable y covers all possible values of the quantity in question. But now it is not to be supposed that all these values are actually to be found in nature. Even if the number of different temperatures which specimens of gases have or will acquire is infinite, there still must be an infinite number missing. How then are we to interpret such a law? As being the compendious assertion of all its actual instances? But the formulation of the law in no way indicates which the actual instances are. It would be absurd to construe a general formula about the functional dependence of one quantity on another as committing us to the assertion that just these values of the quantity are actually realized. As asserting that for a value n of y , which is in fact not realized, the proposition that it is realized, in conjunction with the set of propositions describing all the actual cases, entails the proposition that there is a corresponding value m of x ? But this is open to the same objection, with the further drawback that the entailment would not hold. As asserting with regard to any given value n of y that either n is not realized or that there is a corresponding value m of x ? This is the most plausible alternative, but it makes the law trivial for all the values of y which happen not to be realized. It is hard to escape the conclusion that what we really mean to assert when we formulate such a law is that there is a corresponding value of x to every *possible* value of y .

Another reason for bringing in possibilities is that there seems to be no other way of accounting for the difference between generalizations of law and generalizations of fact. To revert to our earlier examples, it is a generalization of fact that all the Presidents of the Third French Republic are male, or that all the cigarettes that are now in my cigarette case are

made of Virginian tobacco. It is a generalization of law that the planets of our solar system move in elliptical orbits, but a generalization of fact that, counting the earth as Terra, they all have Latin names. Some philosophers refer to these generalizations of fact as 'accidental generalizations', but this use of the word 'accidental' may be misleading. It is not suggested that these generalizations are true by accident, in the sense that there is no causal explanation of their truth, but only that they are not themselves the expression of natural laws.

But how is this distinction to be made? The formula $(x)(\Phi x \supset \Psi x)$ holds equally in both cases. Whether the generalization be one of fact or of law, it will state at least that there is nothing which has the property Φ but lacks the property Ψ . In this sense, the generality is perfect in both cases, so long as the statements are true. Yet there seems to be a sense in which the generality of what we are calling generalizations of fact is less complete. They seem to be restricted in a way that generalizations of law are not. Either they involve some spatio-temporal restriction, as in the example of the cigarettes now in my cigarette case, or they refer to particular individuals, as in the example of the presidents of France. When I say that all the planets have Latin names, I am referring definitely to a certain set of individuals, Jupiter, Venus, Mercury, and so on, but when I say that the planets move in elliptical orbits I am referring indefinitely to anything that has the properties that constitute being a planet in this solar system. But it will not do to say that generalizations of fact are simply conjunctions of particular statements, which definitely refer to individuals; for in asserting that the planets have Latin names, I do not individually identify them: I may know that they have Latin names without being able to list them all. Neither can we mark off generalizations of law by insisting that their expression is not to include any reference to specific places or times. For with a little ingenuity, generalizations of fact can always be made to satisfy this condition. Instead of referring to the cigarettes that are now in my cigarette case, I can find out some general property which only these cigarettes happen to possess, say the property of being contained in a cigarette case with such and such markings which is owned at such and such a period of his life by a person of such and such a sort, where the descriptions are so chosen that the description of the person is in fact satisfied only by me and the description of the cigarette case, if I possess more than one of them, only by the one in question. In certain instances these descriptions might have to be rather complicated, but usually they would not: and anyhow the question of complexity is not here at issue. But this means that, with the help of these 'individuating' predicates, generalizations of fact can be expressed in just as universal a form as generalizations of law. And conversely, as Professor Nelson Goodman has pointed out, generalizations of law can themselves be expressed in such a way that they contain a reference to particular individuals, or to specific places and times. For, as he remarks, 'even the hypothesis "All grass is

green" has as an equivalent "All grass in London or elsewhere is green".¹⁰ Admittedly, this assimilation of the two types of statement looks like a dodge; but the fact that the dodge works shows that we cannot found the distinction on a difference in the ways in which the statement can be expressed. Again, what we want to say is that whereas generalizations of fact cover only actual instances, generalizations of law cover possible instances as well. But this notion of possible, as opposed to actual, instances has not yet been made clear.

If generalizations of law do cover possible as well as actual instances, their range must be infinite; for while the number of objects which do throughout the course of time possess a certain property may be finite, there can be no limit to the number of objects which might possibly possess it: for once we enter the realm of possibility we are not confined even to such objects as actually exist. And this shows how far removed these generalizations are from being conjunctions: not simply because their range is infinite, which might be true even if it were confined to actual instances, but because there is something absurd about trying to list all the possible instances. One can imagine an angel's undertaking the task of naming or describing all the men that there ever have been or will be, even if their number were infinite, but how would he set about naming, or describing, all the possible men? This point is developed by F. P. Ramsey who remarks that the variable hypothetical $(x)\Phi x$ resembles a conjunction (a) in that it contains all lesser, *i.e.* here all finite conjunctions, and appears as a sort of infinite product. (b) When we ask what would make it true, we inevitably answer that it is true if and only if every x has Φ ; *i.e.* when we regard it as a proposition capable of the two cases truth and falsity, we are forced to make it a conjunction which we cannot express for lack of symbolic power'.¹¹ But, he goes on, 'what we can't say we can't say, and we can't whistle it either', and he concludes that the variable hypothetical is not a conjunction and that 'if it is not a conjunction, it is not a proposition at all'. Similarly, Professor Ryle, without explicitly denying that generalizations of law are propositions, describes them as 'seasonal inference warrants',¹² on the analogy of season railway-tickets, which implies that they are not so much propositions as rules. Professor Schlick also held that they were rules, arguing that they could not be propositions because they were not conclusively verifiable; but this is a poor argument, since it is doubtful if any propositions are conclusively verifiable, except possibly those that describe the subject's immediate experiences.

Now to say that generalizations of law are not propositions does have the merit of bringing out their peculiarity. It is one way of emphasizing the difference between them and generalizations of fact. But I think that it emphasizes it too strongly. After all, as Ramsey himself acknowledges, we do want to say that generalizations of law are either true or false. And

they are tested in the way that other propositions are, by the examination of actual instances. A contrary instance refutes a generalization of law in the same way as it refutes a generalization of fact. A positive instance confirms them both. Admittedly, there is the difference that if all the actual instances are favourable, their conjunction entails the generalization of fact, whereas it does not entail the generalization of law: but still there is no better way of confirming a generalization of law than by finding favourable instances. To say that lawlike statements function as seasonal inference warrants is indeed illuminating, but what it comes to is that the inferences in question are warranted by the facts. There would be no point in issuing season tickets if the trains did not actually run.

To say that generalizations of law cover possible as well as actual cases is to say that they entail subjunctive conditionals. If it is a law of nature that the planets move in elliptical orbits, then it must not only be true that the actual planets move in elliptical orbits; it must also be true that if anything were a planet it would move in an elliptical orbit: and here 'being a planet' must be construed as a matter of having certain properties, not just as being identical with one of the planets that there are. It is not indeed a peculiarity of statements which one takes as expressing laws of nature that they entail subjunctive conditionals: for the same will be true of any statement that contains a dispositional predicate. To say, for example, that this rubber band is elastic is to say not merely that it will resume its normal size when it has been stretched, but that it would do so if ever it were stretched: an object may be elastic without ever in fact being stretched at all. Even the statement that this is a white piece of paper may be taken as implying not only how the piece of paper does look but also how it would look under certain conditions, which may or may not be fulfilled. Thus one cannot say that generalizations of fact do not entail subjunctive conditionals, for they may very well contain dispositional predicates: indeed they are more likely to do so than not: but they will not entail the subjunctive conditionals which are entailed by the corresponding statements of law. To say that all the planets have Latin names may be to make a dispositional statement, in the sense that it implies not so much that people do always call them by such names but that they would so call them if they were speaking correctly. It does not, however, imply with regard to anything whatsoever that if it were a planet it would be called by a Latin name. And for this reason it is not a generalization of law, but only a generalization of fact.

There are many philosophers who are content to leave the matter there. They explain the 'necessity' of natural laws as consisting in the fact that they hold for all possible, as well as actual, instances: and they distinguish generalizations of law from generalizations of fact by bringing out the differences in their entailment of subjunctive conditionals. But while this is correct so far as it goes, I doubt if it goes far enough. Neither the

notion of possible, as opposed to actual, instances nor that of the subjunctive conditional is so pellucid that these references to them can be regarded as bringing all our difficulties to an end. It will be well to try to take our analysis a little further if we can.

The theory which I am going to sketch will not avoid all talk of dispositions; but it will confine it to people's attitudes. My suggestion is that the difference between our two types of generalization lies not so much on the side of the facts which make them true or false, as in the attitude of those who put them forward. The factual information which is expressed by a statement of the form 'for all x , if x has Φ then x has Ψ ', is the same whichever way it is interpreted. For if the two interpretations differ only with respect to the possible, as opposed to the actual values of x , they do not differ with respect to anything that actually happens. Now I do not wish to say that a difference in regard to mere possibilities is not a genuine difference, or that it is to be equated with a difference in the attitude of those who do the interpreting. But I do think that it can best be elucidated by referring to such differences of attitude. In short I propose to explain the distinction between generalizations of law and generalizations of fact, and thereby to give some account of what a law of nature is, by the indirect method of analysing the distinction between treating a generalization as a statement of law and treating it as a statement of fact.

If someone accepts a statement of the form ' $(x)(\Phi x \supset \Psi x)$ ' as a true generalization of fact, he will not in fact believe that anything which has the property Φ has any other property that leads to its not having Ψ . For since he believes that everything that has Φ has Ψ , he must believe that whatever other properties a given value of x may have they are not such as to prevent its having Ψ . It may be even that he knows this to be so. But now let us suppose that he believes such a generalization to be true, without knowing it for certain. In that case there will be various properties X, X_1, \dots such that if he were to learn, with respect to any value of a of x , that a had one or more of these properties as well as Φ , it would destroy, or seriously weaken his belief that a had Ψ . Thus I believe that all the cigarettes in my case are made of Virginian tobacco, but this belief would be destroyed if I were informed that I had absent-mindedly just filled my case from a box in which I keep only Turkish cigarettes. On the other hand, if I took it to be a law of nature that all the cigarettes in this case were made of Virginian tobacco, say on the ground that the case had some curious physical property which had the effect of changing any other tobacco that was put into it into Virginian, then my belief would not be weakened in this way.

Now if our laws of nature were causally independent of each other, and if, as Mill thought, the propositions which expressed them were always put forward as being unconditionally true, the analysis could proceed quite simply. We could then say that a person A was treating a statement of the

form 'for all x , if Φx then Ψx ' as expressing a law of nature, if and only if there was no property X which was such that the information that a value a of x had X as well as Φ would weaken his belief that a had Ψ . And here we should have to admit the proviso that X did not logically entail not- Ψ , and also, I suppose, that its presence was not regarded as a manifestation of not- Ψ ; for we do not wish to make it incompatible with treating a statement as the expression of a law that one should acknowledge a negative instance if it arises. But the actual position is not so simple. For one may believe that a statement of the form 'for all x , if Φx then Ψx ' expresses a law of nature while also believing, because of one's belief in other laws, that if something were to have the property X as well as Φ it would not have Ψ . Thus one's belief in the proposition that an object which one took to be a loadstone attracted iron might be weakened or destroyed by the information that the physical composition of the supposed loadstone was very different from what one had thought it to be. I think, however, that in all such cases, the information which would impair one's belief that the object in question had the property Ψ would also be such that, independently of other considerations, it would seriously weaken one's belief that the object ever had the property Φ . And if this is so, we can meet the difficulty by stipulating that the range of properties which someone who treats 'for all x , if Φx then Ψx ' as a law must be willing to conjoin with Φ , without his belief in the consequent being weakened, must not include those the knowledge of whose presence would in itself seriously weaken his belief in the presence of Φ .

There remains the further difficulty that we do not normally regard the propositions which we take to express laws of nature as being unconditionally true. In stating them we imply the presence of certain conditions which we do not actually specify. Perhaps we could specify them if we chose, though we might find it difficult to make the list exhaustive. In this sense a generalization of law may be weaker than a generalization of fact, since it may admit exceptions to the generalization as it is stated. This does not mean, however, that the law allows for exceptions: if the exception is acknowledged to be genuine, the law is held to be refuted. What happens in the other cases is that the exception is regarded as having been tacitly provided for. We lay down a law about the boiling point of water, without bothering to mention that it does not hold for high altitudes. When this is pointed out to us, we say that this qualification was meant to be understood. And so in other instances. The statement that if anything has Φ it has Ψ was a loose formulation of the law: what we really meant was that if anything has Φ but not X , it has Ψ . Even in the case where the existence of the exception was not previously known, we often regard it as qualifying rather than refuting the law. We say, not that the generalization has been falsified, but that it was inexactly stated. Thus, it must be allowed that someone whose belief in the presence of Ψ , in a given

instance, is destroyed by the belief that Φ is accompanied by X may still be treating ' $(x)(\Phi x \supset \Psi x)$ ' as expressing a law of nature if he is prepared to accept ' $(x)((\Phi x \cdot \sim Xx) \supset \Psi x)$ ' as a more exact statement of the law.

Accordingly I suggest that for someone to treat a statement of the form 'if anything has Φ it has Ψ ' as expressing a law of nature, it is sufficient (i) that subject to a willingness to explain away exceptions he believes that in a non-trivial sense everything which in fact has Φ has Ψ (ii) that his belief that something which has Φ has Ψ is not liable to be weakened by the discovery that the object in question also has some other property X , provided (a) that X does not logically entail not- Ψ (b) that X is not a manifestation of not- Ψ (c) that the discovery that something had X would not in itself seriously weaken his belief that it had Φ (d) that he does not regard the statement 'if anything has Φ and not- X it has Ψ ' as a more exact statement of the generalization that he was intending to express.

I do not suggest that these conditions are necessary, both because I think it possible that they could be simplified and because they do not cover the whole field. For instance, no provision has been made for functional laws, where the reference to possible instances does not at present seem to me eliminable. Neither am I offering a definition of natural law. I do not claim that to say that some proposition expresses a law of nature entails saying that someone has a certain attitude towards it; for clearly it makes sense to say that there are laws of nature which remain unknown. But this is consistent with holding that the notion is to be explained in terms of people's attitudes. My explanation is indeed sketchy, but I think that the distinctions which I have tried to bring out are relevant and important: and I hope that I have done something towards making them clear.

■ | Notes

1. *Leviathan*, Part I, Chap. xv.
2. *An Enquiry Concerning Human Understanding*, iv, I.25.
3. *Probability and Induction*, pp. 79 ff.
4. *A Treatise of Human Nature*, i, iii, vi.
5. Cf. *La Science et l'hypothèse* [Science and hypothesis], pp. 119–29.
6. Cf. Kneale, *op. cit.*
7. Cf. K. Popper, "What Can Logic Do for Philosophy?" *Supplementary Proceedings of the Aristotelian Society*, Vol. XXII: and papers in the same volume by W. C. Kneale and myself.
8. *Scientific Explanation*, pp. 118–29.

9. "Mechanical and Teleological Causation," *Supplementary Proceedings of the Aristotelian Society*, XIV, 98 ff.
10. *Fact, Fiction and Forecast*, p. 78.
11. *Foundations of Mathematics*, p. 238.
12. "'If,' 'So,' and 'Because'," *Philosophical Analysis* (Essays edited by Max Black), p. 332.

Laws of Nature

It is tempting to identify the laws of nature with a certain class of universal truths. Very few empiricists have succeeded in resisting this temptation. The popular way of succumbing is to equate the fundamental laws of nature with what is asserted by those universally true statements of non-limited scope that embody only qualitative predicates.¹ On this view of things a law-like statement is a statement of the form " $(x)(Fx \supset Gx)$ " or " $(x)(Fx \equiv Gx)$ " where " F " and " G " are purely qualitative (nonpositional). Those law-like statements that are true express laws. "All robins' eggs are greenish blue," "All metals conduct electricity," and "At constant pressure any gas expands with increasing temperature" (Hempel's examples) are law-like statements. If they are true, they express laws. The more familiar sorts of things that we are accustomed to calling laws, the formulae and equations appearing in our physics and chemistry books, can supposedly be understood in the same way by using functors in place of the propositional functions " Fx " and " Gx " in the symbolic expressions given above.*

I say that it is tempting to proceed in this way since, to put it bluntly,

FROM *Philosophy of Science* 44 (1977): 248–68.

* Although it does not affect any of the philosophical issues debated in this chapter, Dretske's remark about functors raises an interesting question, namely, whether predicate logic has the resources to represent quantitative laws adequately. It seems most unlikely that a so-called functional law written as an equation involving several variables, each of which takes real numbers as values, can be properly regarded as having the simple form $(x)(Fx \supset Gx)$, where F and G are qualitative predicates. The functors mentioned by Dretske attempt to solve this problem by converting the equation into a function that is then treated as a predicate. Consider the ideal gas law, $PV = nRT$. The functor in this case might be, "is identical with the value of nRT divided by V " and the law would read (roughly): "For all x , if x is the value of the pressure of an ideal gas, then x is identical with the value of nRT divided by V ." Not only is this clumsy but, by focusing on pressure in the antecedent, it obscures the interdependence of the variables: the ideal gas law is not about pressure; it is about all the variables and their functional relation. Alternatively, we might define the predicate "obeys the equation $PV = nRT$ " and then portray the ideal gas law as: "For all x , if x is an ideal gas, then x obeys the equation $PV = nRT$." But this says merely that all ideal gases obey the ideal gas law, which can hardly be regarded as a perspicuous representation of the law.

conceiving of a law as having a content greater than that expressed by a statement of the form $(x)(Fx \supset Gx)$ seems to put it beyond our epistemological grasp.² We must work with what we are given, and what we are given (the observational and experimental data) are facts of the form: this F is G , that F is G , all examined F 's have been G , and so on. If, as some philosophers have argued,³ law-like statements express a kind of nomic necessity between events, something *more* than that F 's are, as a matter of fact, always and everywhere, G , then it is hard to see what kind of evidence might be brought in support of them. The whole point in acquiring instantial evidence (evidence of the form "This F is G ") in support of a law-like hypothesis would be lost if we supposed that what the hypothesis was actually asserting was some kind of nomic connection, some kind of modal relationship, between things that were F and things that were G . We would, it seems, be in the position of someone trying to confirm the *analyticity* of "All bachelors are unmarried" by collecting evidence about the marital status of various bachelors. This kind of evidence, though relevant to the *truth* of the claim that all bachelors are unmarried, is powerless to confirm the *modality* in question. Similarly, if a hypothesis, in order to qualify as a law, must express or assert some form of necessity between F 's and G 's, then it becomes a mystery how we ever manage to confirm such attributions with the sort of instantial evidence available from observation.

Despite this argument, the fact remains that laws are *not* simply what universally true statements express, not even universally true statements that embody purely qualitative predicates (and are, as a result, unlimited in scope). This is not particularly newsworthy. It is commonly acknowledged that law-like statements have some peculiarities that prevent their straightforward assimilation to universal truths. That the concept of a law and the concept of a universal truth are different concepts can best be seen, I think, by the following consideration: assume that $(x)(Fx \supset Gx)$ is true and that the predicate expressions satisfy all the restrictions that one might wish to impose in order to convert this universal statement into a statement of law.⁴ Consider a predicate expression " K " (eternally) coextensive with " F "; i.e., $(x)(Fx = Kx)$ for all time. We may then infer that if $(x)(Fx \supset Gx)$ is a universal truth, so is $(x)(Kx \supset Gx)$. The class of universal truths is closed under the operation of coextensive predicate substitution. Such is *not* the case with laws. If it is a law that all F 's are G , and we substitute the term " K " for the term " F " in this law, the result is not necessarily a law. If diamonds have a refractive index of 2.419 (law) and "is a diamond" is coextensive with "is mined in kimberlite (a dark basic rock)" we cannot infer that *it is a law* that things mined in kimberlite have a refractive index of 2.419. Whether this is a law or not depends on whether the coextensiveness of "is a diamond" and "is mined in kimberlite" is *itself* law-like. The class of laws is not closed under the same operation as is the class of universal truths.

Using familiar terminology we may say that the predicate positions in a statement of law are *opaque* while the predicate positions in a universal truth of the form $(x)(Fx \supset Gx)$ are *transparent*.^{*} I am using these terms in a slightly unorthodox way. It is not that when we have a law, "All *F*'s are *G*," we can alter its truth value by substituting a coextensive predicate for "*F*" or "*G*." For if the statement is true, it will remain true after substitution. What happens, rather, is that the expression's status as a law is (or may be) affected by such an exchange. The matter can be put this way: the statement

(A) All *F*'s are *G* (understood as $(x)(Fx \supset Gx)$)

has "*F*" and "*G*" occurring in transparent positions. Its truth value is unaffected by the replacement of "*F*" or "*G*" by a coextensive predicate. The same is true of

(B) It is universally true that *F*'s are *G*.

If, however, we look at

(C) It is a law that *F*'s are *G*.

we find that "*F*" and "*G*" occur in opaque positions. If we think of the two prefixes in (B) and (C), "it is universally true that . . ." and "it is a law that . . .," as operators, we can say that the operator in (B) does not, while the operator in (C) does, confer opacity on the embedded predicate positions. To refer to something as a statement of law is to refer to it as an expression in which the descriptive terms occupy opaque positions. To refer to something as a universal truth is to refer to it as an expression in

^{*} *Transparent* and *opaque* are terms used in the theory of reference. Consider the true sentence, "Blue whales live in water." If we replace the expression "blue whales" with a phrase that designates the same class of animals—"the largest mammals on earth," for example—then the sentence must remain true. Philosophers of language say that "Blue whales live in water" is a transparent context because its truth value cannot be altered by the substitution of coreferring expressions. Contrast this with "John knows that blue whales live in water." This is an opaque context. John might be ignorant of the fact that blue whales are the largest mammals on earth. Hence, he could know that blue whales live in water without also knowing that the largest mammals on earth live in water. Thus, in this case, the substitution of a coreferring expression could change a true sentence into a false one. Modal contexts (that is, sentences involving possibility or necessity) can also create opacity. "Blue whales are necessarily whales" is true but "The largest mammals on earth are necessarily whales" is false because it is contingent, not necessary, that the world's largest mammals happen to be whales. Similarly, Dretske argues, sentences of the form "It is a law that all *F*s are *G*" are referentially opaque.

which the descriptive terms occupy transparent positions. Hence, our concept of a law differs from our concept of a universal truth.⁵

Confronted by a difference of this sort, many philosophers have argued that the distinction between a natural law and a universal truth was not, fundamentally, an *intrinsic* difference. Rather, the difference was a difference in the *role* some universal statements played within the larger theoretical enterprise. Some universal statements are more highly integrated into the constellation of accepted scientific principles, they play a more significant role in the explanation and prediction of experimental results, they are better confirmed, have survived more tests, and make a more substantial contribution to the regulation of experimental inquiry. But, divorced from this context, stripped of these *extrinsic* features, a law is nothing but a universal truth. It has the same empirical content. Laws are to universal truths what shims are to slivers of wood and metal; the latter *become* the former by being *used* in a certain way. There is a *functional* difference, nothing else.⁶

According to this reductionistic view, the peculiar opacity (described above) associated with laws is not a manifestation of some intrinsic difference between a law and a universal truth. It is merely a symptom of the special status or function that some universal statements have. The basic formula is: law = universal truth + X. The "X" is intended to indicate the special function, status or role that a universal truth must have to *qualify* as a law. Some popular candidates for this auxiliary idea, X, are:

- 1 High degree of confirmation,
- 2 Wide acceptance (well established in the relevant community),
- 3 Explanatory potential (can be used to explain its instances),
- 4 Deductive integration (within a larger system of statements),
- 5 Predictive use.

To illustrate the way these values of X are used to buttress the equation of laws with universal truths, it should be noted that each of the concepts appearing on this list generates an opacity similar to that witnessed in the case of genuine laws. For example, to say that it is a law that all F's are G may possibly be no more than to say that it is well established that $(x)(Fx \supset Gx)$. The peculiar opacity of laws is then explained by pointing out that the class of expressions that are well established (or highly confirmed) is not closed under substitution of coextensive predicates: one cannot infer that $(x)(Kx \supset Gx)$ is well established just because "Fx" and "Kx" are coextensive and $(x)(Fx \supset Gx)$ is well established (for no one may know that "Fx" and "Kx" are coextensive). It may be supposed, therefore, that the opacity of laws is merely a manifestation of the underlying fact that a universal statement, to qualify as a law, must be well established, and the opacity is a result of this epistemic condition. Or, if this will not

do, we can suppose that one of the other notions mentioned above, or a combination of them, is the source of a law's opacity.

This response to the alleged uniqueness of natural laws is more or less standard fare among empiricists in the Humean tradition. Longstanding (= venerable) epistemological and ontological commitments motivate the equation: law = universal truth + X. There is disagreement among authors about the differentia X, but there is near unanimity about the fact that laws are a *species* of universal truth.

If we set aside our scruples for the moment, however, there is a plausible explanation for the opacity of laws that has not yet been mentioned. Taking our cue from Frege, it may be argued that since the operator "it is a law that . . ." converts the otherwise transparent positions of "All *F*'s are *G*" into opaque positions, we may conclude that this occurs because within the context of this operator (either explicitly present or implicitly understood) the terms "*F*" and "*G*" do not have their usual referents. There is a shift in what we are talking about. To say that *it is a law* that *F*'s are *G* is to say that "All *F*'s are *G*" is to be understood (in so far as it expresses a law), not as a statement about the extensions of the predicates "*F*" and "*G*," but as a singular statement describing a relationship between the universal properties *F*-ness and *G*-ness. In other words, (C) is to be understood as having the form:

$$6 \text{ } F\text{-ness} \rightarrow G\text{-ness.}^7$$

To conceive of (A) as a universal truth is to conceive of it as expressing a relationship between the extensions of its terms; to conceive of it as a law is to conceive of it as expressing a relationship between the properties (magnitudes, quantities, features) which these predicates express (and to which we may refer with the corresponding abstract singular term). The opacity of laws is merely a manifestation of this change in reference. If "*F*" and "*K*" are coextensive, we cannot substitute the one for the other in the law "All *F*'s are *G*" and expect to preserve truth; for the law asserts a connection between *F*-ness and *G*-ness and there is no guarantee that a similar connection exists between the properties *K*-ness and *G*-ness just because all *F*'s are *K* and *vice versa*.⁸

It is this view that I mean to defend in the remainder of this essay. Law-like statements are singular statements of fact describing a relationship between properties or magnitudes. Laws are the relationships that are asserted to exist by true law-like statements. According to this view, then, there is an *intrinsic* difference between laws and universal truths. Laws imply universal truths, but universal truths do not imply laws. Laws are (expressed by) *singular* statements describing the relationships that exist between universal qualities and quantities; they are not universal statements about the particular objects and situations that exemplify these qualities and quantities. Universal truths are not transformed into laws by

acquiring some of the extrinsic properties of laws, by being used in explanation or prediction, by being made to support counterfactuals, or by becoming well established. For, as we shall see, universal truths *cannot* function in these ways. They *cannot* be made to perform a service they are wholly unequipped to provide.

In order to develop this thesis it will be necessary to overcome some metaphysical prejudices, and to overcome these prejudices it will prove useful to review the major deficiencies of the proposed alternative. The attractiveness of the formula: law = universal truth + X, lies, partly at least, in its ontological austerity, in its tidy portrayal of what there is, or what there must be, in order for there to be laws of nature. The antidote to this seductive doctrine is a clear realization of how utterly hopeless, epistemologically and functionally hopeless, this equation is.

If the auxiliary ideas mentioned above (explanation, prediction, confirmation, etc.) are deployed as values of X in the reductionistic equation of laws with universal truths, one can, as we have already seen, render a satisfactory account of the opacity of laws. In this particular respect the attempted equation proves adequate. In what way, then, does it fail?

(1) and (2) are what I will call "epistemic" notions; they assign to a statement a certain epistemological status or cognitive value. They are, for this reason alone, useless in understanding the nature of a law.⁹ Laws do not begin to be laws only when we first become aware of them, when the relevant hypotheses become well established, when there is public endorsement by the relevant scientific community. The laws of nature are the same today as they were one thousand years ago (or so we believe); yet, some hypotheses are highly confirmed today that were not highly confirmed one thousand years ago. It is certainly true that we only begin to *call* something a law when it becomes well established, that we only recognize something as a statement of law when it is confirmed to a certain degree, but that something is a law, that some statement does in fact express a law, does not similarly await our appreciation of this fact. We discover laws, we do not invent them—although, of course, some invention may be involved in our manner of expressing or codifying these laws. Hence, the status of something as a statement of law does not depend on its epistemological status. What does depend on such epistemological factors is our ability to identify an otherwise qualified statement as *true* and, therefore, as a *statement of law*. It is for this reason that one cannot appeal to the epistemic operators to clarify the nature of laws; they merely confuse an epistemological with an ontological issue.

What sometimes helps to obscure this point is the tendency to conflate laws with the verbal or symbolic expression of these laws (what I have been calling "statements of law"). Clearly, though, these are different things and should not be confused. There are doubtless laws that have not yet (or will never) receive symbolic expression, and the same law may be given different verbal codifications (think of the variety of ways of express-

ing the laws of thermodynamics). To use the language of "propositions" for a moment, a law is the proposition expressed, not the vehicle we use to express it. The *use* of a sentence *as an expression of law* depends on epistemological considerations, but the law itself does not.

There is, furthermore, the fact that whatever auxiliary idea we select for understanding laws (as candidates for X in the equation: law = universal truth $\rightarrow X$), if it is going to achieve what we expect of it, should help to account for the variety of other features that laws are acknowledged to have. For example, it is said that laws "support" counterfactuals of a certain sort. If laws are universal truths, this fact is a complete mystery, a mystery that is usually suppressed by using the word "support." For, of course, universal statements do not *imply* counterfactuals in any sense of the word "imply" with which I am familiar. To be told that all F 's are G is not to be told anything that implies that if this x were an F , it would be G . To be told that all dogs born at sea have been and will be cocker spaniels is *not* to be told that we would get cocker spaniel pups (or no pups at all) if we arranged to breed dachshunds at sea. The only reason we might *think* we were being told this is because we do not expect anyone to assert that all dogs born at sea *will be* cocker spaniels unless they know (or have good reasons for believing) that this is true; and we do not understand *how* anyone could *know* that this is true without being privy to information that insures this result—without, that is, knowing of some bizzare law or circumstance that *prevents* anything but cocker spaniels from being born at sea. Hence, *if* we accept the claim at all, we do so with a certain presumption about what our informant must know in order to be a serious claimant. We assume that our informant knows of certain laws or conditions that *insure* the continuance of a past regularity, and it is this presumed knowledge that we exploit in endorsing or accepting the counterfactual. But the simple fact remains that the statement "All dogs born at sea have been and will be cocker spaniels" does not *itself* support or imply this counterfactual; at best, *we* support the counterfactual (if we support it at all) on the basis of what the claimant is supposed to know in order to advance such a universal projection.

Given this incapacity on the part of universal truths to support counterfactuals, one would expect some assistance from the epistemic condition if laws are to be analyzed as well established universal truths. But the expectation is disappointed; we are *left* with a complete mystery. For if a statement of the form "All F 's are G " does not support the counterfactual, "If this (non- G) were an F , it would be G ," it is clear that it will not support it just because it is well established or highly confirmed. The fact that all the marbles in the bag are red does not support the contention that if this (blue) marble were in the bag, it would be red; but neither does the fact that we *know* (or it is highly confirmed) that all the marbles in the bag are red support the claim that if this marble were in the bag it would be red. And making the universal truth *more universal* is not going

to repair the difficulty. The fact that all the marbles in the universe are (have been and will be) red does not imply that I *cannot* manufacture a blue marble; it implies that I *will not*, not that I cannot or that if I were to try, I would fail. To represent laws on the model of one of our epistemic operators, therefore, leaves wholly unexplained one of the most important features of laws that we are trying to understand. They are, in this respect, unsatisfactory candidates for the job.

Though laws are not merely well established general truths, there is a related point that deserves mention: laws are the *sort* of thing that can become well established prior to an exhaustive enumeration of the instances to which they apply. This, of course, is what gives laws their predictive utility. Our confidence in them increases at a much more rapid rate than does the ratio of favorable examined cases to total number of cases. Hence, we reach the point of confidently using them to project the outcome of unexamined situations while there is still a substantial number of unexamined situations to project.

This feature of laws raises new problems for the reductionistic equation. For, contrary to the argument in the second paragraph of this essay, it is hard to see how confirmation is possible for universal truths. To illustrate this difficulty, consider the (presumably easier) case of a general truth of *finite* scope. I have a coin that you have (by examination and test) convinced yourself is quite normal. I propose to flip it ten times. I conjecture (for whatever reason) that it will land heads all ten times. You express doubts. I proceed to "confirm" my hypothesis. I flip the coin once. It lands heads. Is this evidence that my hypothesis is correct? I continue flipping the coin and it turns up with nine straight heads. Given the opening assumption that we are dealing with a fair coin, the probability of getting all ten heads (the probability that my hypothesis is true) is now, after examination of 90% of the total population to which the hypothesis applies, exactly .5. If we are guided by probability considerations alone, the likelihood of all ten tosses being heads is now, after nine favorable trials, a toss-up. After nine favorable trials it is no more reasonable to believe the hypothesis than its denial. In what sense, then, can we be said to have been accumulating evidence (during the first nine trials) that all would be heads? In what sense have we been confirming the hypothesis? It would appear that the probability of my conjecture's being true never exceeds .5 until we have exhaustively examined the entire population of coin tosses and found them *all* favorable. The probability of my conjecture's being true is either: (i) too low ($\leq .5$) to invest any confidence in the hypothesis, or (ii) so high ($= 1$) that the hypothesis is useless for prediction. There does not seem to be any middle ground.

Our attempts to confirm universal generalizations of nonlimited scope is, I submit, in exactly the same impossible situation. It is true, of course, that after nine successful trials the probability that all ten tosses will be heads is greatly increased over the initial probability that all would be

heads. The initial probability (assuming a fair coin) that all ten tosses would be heads was on the order of .002. After nine favorable trials it is .5. In this sense I have increased the probability that my hypothesis is true; I have raised its probability from .002 to .5. The important point to notice, however, is that this sequence of trials did not alter the probability that the *tenth* trial would be heads. The probability that the unexamined instance would be favorable remains exactly what it was before I began flipping the coin. It was originally .5 and it is now, after nine favorable trials, still .5. I am in no better position now, after extensive sampling, to predict the outcome of the tenth toss than I was before I started. To suppose otherwise is to commit the converse of the Gambler's Fallacy.

Notice, we could take the first nine trials as evidence that the tenth trial would be heads *if* we took the results of the first nine tosses as evidence that the coin was biased in some way. Then, on *this* hypothesis, the probability of getting heads on the last trial (and, hence, on all ten trials) would be greater than .5 (how much greater would depend on the conjectured degree of bias and this, in turn, would presumably depend on the extent of sampling). This new hypothesis, however, is something quite different than the original one. The original hypothesis was of the form: $(x)(Fx \supset Gx)$, all ten tosses will be heads. Our new conjecture is that there is a physical asymmetry in the coin, an asymmetry that tends to yield more heads than tails. We have succeeded in confirming the general hypothesis (all ten tosses will be heads), but we have done so via an intermediate hypothesis involving *genuine laws* relating the physical make-up of the coin to the frequency of heads in a population of tosses.

It is by such devices as this that we create for ourselves, or some philosophers create for themselves, the *illusion* that (apart from supplementary *law-like* assumptions) general truths can be confirmed by their instances and therefore qualify, in this respect, as laws of nature. The illusion is fostered in the following way. It is assumed that confirmation is a matter of *raising the probability of a hypothesis*.¹⁰ On this assumption any general statement of finite scope can be confirmed by examining its instances and finding them favorable. The hypothesis about the results of flipping a coin ten times can be confirmed by tossing nine straight heads, and this confirmation takes place without *any* assumptions about the coin's bias. Similarly, I confirm (to some degree) the hypothesis that all the people in the hotel ballroom are over thirty years old when I enter the ballroom with my wife and realize that *we* are both over thirty. In both cases I raise the probability that the hypothesis is true over what it was originally (before flipping the coin and before entering the ballroom). But this, of course, isn't confirmation. Confirmation is not simply raising the probability that a hypothesis is true, it is raising the probability that the unexamined cases resemble (in the relevant respect) the examined cases. It is *this* probability that must be raised if genuine confirmation is to occur (and if a confirmed hypothesis is to be useful in *prediction*), and it is precisely

this probability that is left unaffected by the instantial "evidence" in the above examples.

In order to meet this difficulty, and to cope with hypotheses that are *not* of limited scope,¹¹ the reductionist usually smuggles into his confirmatory proceedings the very idea he professes to do without: *viz.*, a type of law that is not merely a universal truth. The general truth then gets confirmed but *only* through the mediation of these supplementary laws. These supplementary assumptions are usually introduced to *explain* the regularities manifested in the examined instances so as to provide a basis for projecting these regularities to the unexamined cases. The only way we can get a purchase on the unexamined cases is to introduce a hypothesis which, while *explaining* the data we already have, *implies* something about the data we do not have. To suppose that our coin is biased (first example) is to suppose something that contributes to the explanation of our extraordinary run of heads (nine straight) and simultaneously implies something about the (probable) outcome of the tenth toss. Similarly (second example) my wife and I may be attending a reunion of some kind, and I may suppose that the other people in the ballroom are old classmates. This hypothesis not only explains our presence, it implies that most, if not all, of the remaining people in the room are of comparable age (well over thirty). In both these cases the generalization can be confirmed, but only via the introduction of a law or circumstance (combined with a law or laws) that helps to explain the data already available.

One additional example should help to clarify these last remarks. In sampling from an urn with a population of colored marbles, I can confirm the hypothesis that all the marbles in the urn are red by extracting at random several dozen red marbles (and no marbles of any other color). This is a genuine example of confirmation, not because I have raised the probability of the hypothesis that all are red by reducing the number of ways it can be false (the same reduction would be achieved if you *showed* me 24 marbles from the urn, all of which were red), but because the hypothesis that all the marbles in the urn are red, together with the fact (law) that you cannot draw nonred marbles from an urn containing only red marbles, *explains* the result of my random sampling. Or, if this is too strong, the law that assures me that random sampling from an urn containing a substantial number of nonred marbles would reveal (in all likelihood) at least one nonred marble lends its support to my confirmation that the urn contains only (or mostly) red marbles. Without the assistance of such auxiliary laws a sample of 24 red marbles is powerless to confirm a hypothesis about the total population of marbles in the urn. To suppose otherwise is to suppose that the *same* degree of confirmation would be afforded the hypothesis if you, whatever your deceitful intentions, showed me a carefully selected set of 24 red marbles from the urn. This *also* raises the probability that they are all red, but the trouble is that it does not (due to your unknown motives and intentions) raise the probability that the

unexamined marbles resemble the examined ones. And it does not raise this probability because we no longer have, as the best available explanation of the examined cases (all red), a hypothesis that implies that the remaining (or most of the remaining) marbles are also red. Your careful selection of 24 red marbles from an urn containing many different colored marbles is an equally good explanation of the data and it does *not* imply that the remainder are red. Hence, it is not just the fact that we have 24 red marbles in our sample class (24 positive instances and no negative instances) that confirms the general hypothesis that all the marbles in the urn are red. It is this data *together with a law* that confirms it, a law that (together with the hypothesis) explains the data in a way that the general hypothesis alone cannot do.

We have now reached a critical stage in our examination of the view that a properly qualified set of universal generalizations can serve as the fundamental laws of nature. For we have, in the past few paragraphs, introduced the notion of *explanation*, and it is this notion, perhaps more than any other, that has received the greatest attention from philosophers in their quest for the appropriate X in the formula: law = universal truth + X. R. B. Braithwaite's treatment ([3]) is typical. He begins by suggesting that it is merely deductive integration that transforms a universal truth into a law of nature. Laws are simply universally true statements of the form $(x)(Fx \supset Gx)$ that are derivable from certain higher level hypotheses. To say that $(x)(Fx \supset Gx)$ is a statement of law is to say, not only that it is true, but that it is *deducible* from a higher level hypothesis, *H*, in a well established scientific system. The fact that it must be deducible from some higher level hypothesis, *H*, confers on the statement the opacity we are seeking to understand. For we may have a hypothesis from which we can derive $(x)(Fx \supset Gx)$ but from which we cannot derive $(x)(Kx \supset Gx)$ despite the coextensionality of "F" and "K." Braithwaite also argues that such a view gives a satisfactory account of the counterfactual force of laws.

The difficulty with this approach (a difficulty that Braithwaite recognizes) is that it only postpones the problem. Something is not a statement of law simply because it is true and deducible from some well-established higher level hypothesis. For every generalization implies another of smaller scope (e.g. $(x)(Fx \supset Gx)$ implies $(x)(Fx \cdot Hx \supset Gx)$), but this fact has not the slightest tendency to transform the latter generalization into a law.* What is required is that the higher level hypothesis *itself* be law-like. You cannot give to others what you do not have yourself. But now, it seems, we are back where we started from. It is at this point that Braithwaite begins talking about the higher level hypotheses having *explanatory force* with respect to the hypotheses subsumed under them. He is forced into this maneuver to account for the fact that these higher level

* Dretske uses a dot (instead of an ampersand) to stand for *and*.

hypotheses—not themselves law-like on his characterization (since not themselves derivable from still higher level hypotheses)—are capable of conferring lawlikeness on their consequences. The higher level hypotheses are laws because they explain; the lower level hypotheses are laws because they are deducible from laws. This fancy twist smacks of circularity. Nevertheless, it represents a conversion to *explanation* (instead of *deducibility*) as the fundamental feature of laws, and Braithwaite concedes this: “A hypothesis to be regarded as a natural law must be a general proposition which can be thought to *explain* its instances” ([3], p. 303) and, a few lines later, “Generally speaking, however, a true scientific hypothesis will be regarded as a law of nature if it has an explanatory function with regard to lower-level hypotheses or its instances.” Deducibility is set aside as an incidental (but, on a Hempelian model of explanation, an important) facet of the more ultimate idea of explanation.

There is an added attraction to this suggestion. As argued above, it is difficult to see how instantial evidence can serve to confirm a universal generalization of the form: $(x)(Fx \supset Gx)$. If the generalization has an infinite scope, the ratio “examined favorable cases/total number of cases” never increases. If the generalization has a finite scope, or we treat its probability as something other than the above ratio, we may succeed in raising its probability by finite samples, but it is never clear how we succeed in raising the probability that the unexamined cases resemble the examined cases without invoking laws as auxiliary assumptions. And this is the very notion we are trying to analyze. To this problem the notion of explanation seems to provide an elegant rescue. If laws are those universal generalizations that explain their instances, then following the lead of a number of current authors (notably Harman ([8], [9]); also see Brody ([4])) we may suppose that universal generalizations can be confirmed because confirmation is (roughly) the converse of explanation; *E* confirms *H* if *H* explains *E*. Some universal generalizations can be confirmed; they are those that explain their instances. Equating laws with universal generalizations having explanatory power therefore achieves a neat economy: we account for the confirmability of laws in terms of the explanatory power of those generalizations to which laws are reduced.

To say that a law is a universal truth having explanatory power is like saying that a chair is a breath of air used to seat people. You cannot make a silk purse out of a sow’s ear, not even a very good sow’s ear; and you cannot *make* a generalization, not even a purely universal generalization, explain its instances. The fact that *every F* is *G* fails to explain why *any F* is *G*, and it fails to explain it, not because its explanatory efforts are too feeble to have attracted our attention, but because the explanatory attempt is never even made. The fact that all men are mortal does not explain why you and I are mortal; it *says* (in the sense of *implies*) that we are mortal, but it does not even suggest *why* this might be so. The fact that all ten tosses will turn up heads is a fact that logically guarantees a head

on the tenth toss, but it is not a fact that explains the outcome of this final toss. On one view of explanation, *nothing* explains it. Subsuming an instance under a universal generalization has exactly as much explanatory power as deriving Q from $P \cdot Q$. None.*

If universal truths of the form $(x)(Fx \supset Gx)$ could be *made* to explain their instances, we might succeed in making them into natural laws. But, as far as I can tell, no one has yet revealed the secret for endowing them with this remarkable power.

This has been a hasty and, in some respects, superficial review of the doctrine that laws are universal truths. Despite its brevity, I think we have touched upon the major difficulties with sustaining the equation: law = universal truth + X (for a variety of different values of "X"). The problems center on the following features of laws:

- a A statement of law has its descriptive terms occurring in opaque positions.
- b The existence of laws does not await our identification of them as laws. In this sense they are objective and independent of epistemic considerations.
- c Laws can be confirmed by their instances and the confirmation of a law raises the probability that the unexamined instances will resemble (in the respect described by the law) the examined instances. In this respect they are useful tools for prediction.
- d Laws are not merely summaries of their instances; typically, they figure in the explanation of the phenomena falling within their scope.
- e Laws (in some sense) "support" counterfactuals; to know a law is to know what would happen if certain conditions were realized.
- f Laws tell us what (in some sense) must happen, not merely what has and will happen (given certain initial conditions).

The conception of laws suggested earlier in this essay, the view that laws are expressed by singular statements of fact describing the relationships between properties and magnitudes, proposes to account for these features of laws in a single, unified, way: (a)–(f) are all manifestations of what might be called "ontological ascent," the shift from talking about individual objects and events, or collections of them, to the quantities and qualities that these objects exemplify. Instead of talking about green and red things, we talk about the *colors* green and red. Instead of talking about gases that have a volume, we talk about the volume (temperature, pressure, entropy) that gases have. Laws eschew reference to the things that have length, charge, capacity, internal energy, momentum, spin, and velocity

* For further criticisms of deductive subsumption as sufficient for explanation, see David-Hillel Ruben, "Arguments, Laws, and Explanation" in chapter 6.

in order to talk about these quantities themselves and to describe *their* relationship to each other.

We have already seen how this conception of laws explains the peculiar opacity of law-like statements. Once we understand that a law-like statement is not a statement about the extensions of its constituent terms, but about the intensions (= the quantities and qualities to which we may refer with the abstract singular form of these terms), then the opacity of laws to *extensional* substitution is natural and expected. Once a law is understood to have the form:

$$6 \quad F\text{-ness} \rightarrow G\text{-ness}$$

the relation in question (the relation expressed by " \rightarrow ") is seen to be an *extensional* relation between *properties* with the terms "*F-ness*" and "*G-ness*" occupying *transparent* positions in (6). Any term referring to the same quality or quantity as "*F-ness*" can be substituted for "*F-ness*" in (6) without affecting its truth or its law-likeness. Coextensive terms (terms referring to the same *quantities* and *qualities*) can be freely exchanged for "*F-ness*" and "*G-ness*" in (6) without jeopardizing its truth value. The tendency to treat laws as some kind of *intensional* relation between extensions, as something of the form $(x)(Fx \boxed{N} \rightarrow Gx)$ (where the connective is some kind of modal connective), is simply a mistaken rendition of the fact that laws are extensional relations between intensions.

Once we make the ontological ascent we can also understand the modal character of laws, the feature described in (e) and (f) above. Although true statements having the form of (6) are not themselves *necessary* truths, nor do they describe a modal relationship between the respective qualities, the contingent relationship between properties that is described imposes a modal quality on the particular events falling within its scope. This *F must be G*. Why? Because *F-ness* is linked to *G-ness*; the one property yields or generates the other in much the way a change in the thermal conductivity of a metal yields a change in its electrical conductivity. The pattern of inference is:

$$\begin{array}{l} \text{I } F\text{-ness} \rightarrow G\text{-ness} \\ \text{This is } F \\ \hline \text{This must be } G. \end{array}$$

This, I suggest, is a valid pattern of inference. It is quite unlike the fallacy committed in (II):

$$\begin{array}{l} \text{II } (x)(Fx \supset Gx) \\ \text{This is } F \\ \hline \text{This must be } G. \end{array}$$

The fallacy here consists in the absorption *into* the conclusion of a modality (entailment) that belongs to the relationship *between* the premises and the conclusion. There is no fallacy in (I), and this, I submit, is the source of the “physical” or “nomic” necessity generated by laws. It is this which explains the power of laws to tell us what *would* happen if we did such-and-such and what *could not* happen whatever we did.

I have no proof for the validity of (I). The best I can do is an analogy. Consider the complex set of legal relationships defining the authority, responsibilities, and powers of the three branches of government in the United States. The executive, the legislative, and the judicial branches of government have, according to these laws, different functions and powers. There is nothing *necessary* about the laws themselves; they could be changed. There is no law that prohibits scrapping all the present laws (including the constitution) and starting over again. Yet, given these laws, it follows that the President *must* consult Congress on certain matters, members of the Supreme Court *cannot* enact laws nor declare war, and members of Congress *must* periodically stand for election. The legal code lays down a set of relationships between the various *offices* of government, and this set of relationships (between the abstract offices) impose legal constraints on the individuals who occupy these offices—constraints that we express with such modal terms as “cannot” and “must.” There are certain things the individuals (and collections of individuals—e.g., the Senate) can and cannot do. *Their* activities are subjected to this modal qualification whereas the framework of laws from which this modality arises is itself modality-free. The President (e.g., Ford) *must* consult the Senate on matter M, but the relationship between the *office* of the President and that *legislative body* we call the Senate that makes Gerald Ford’s action obligatory is not *itself* obligatory. There is no law that says that this relationship between the office of President and the upper house of Congress *must* (legally) endure forever and remain indissoluble.

In matters pertaining to the offices, branches and agencies of government the “can” and “cannot” generated by laws are, of course, legal in character. Nevertheless, I think the analogy revealing. Natural laws may be thought of as a set of relationships that exist between the various “offices” that objects sometimes occupy. Once an object occupies such an office, its activities are constrained by the set of relations connecting that office to other offices and agencies; it *must* do some things, and it *cannot* do other things. In both the legal and the natural context the modality at level n is generated by the set of relationships existing between the entities at level $n + 1$. Without this web of higher order relationships there is nothing to support the attribution of constraints to the entities at a lower level.

To think of statements of law as expressing relationships (such as class inclusion) between the extensions of their terms is like thinking of the legal code as a set of universal imperatives directed to a set of particular

individuals. A law that tells us that the United States President must consult Congress on matters pertaining to *M* is not an imperative issued to Gerald Ford, Richard Nixon, Lyndon Johnson, *et al.* The law tells us something about the duties and obligations attending the *Presidency*; only indirectly does it tell us about the obligations of the Presidents (Gerald Ford, Richard Nixon, *et al.*). It tells us about their obligations in so far as they are occupants of this office. If a law was to be interpreted as of the form: "For all *x*, if *x* is (was or will be) President of the United States, then *x* must (legally) consult Congress on matter *M*," it would be incomprehensible why Sally Bickle, were she to be president, would have to consult Congress on matter *M*. For since Sally Bickle never was, and never will be, President, the law, understood as an imperative applying to *actual* Presidents (past, present and future) does not apply to her. Even if there is a possible world in which she becomes President, this does not make her a member of that class of people to which the law applies; for the law, under this interpretation, is directed to that class of people who become President in *this* world, and Sally is not a member of this class. But we all know, of course, that the law does not apply to individuals, or sets of individuals, in this way; it concerns itself, in part, with the offices that people occupy and only indirectly with individuals in so far as they occupy these offices. And this is why, if Sally Bickle were to become President, if she occupied this office, she would have to consult Congress on matters pertaining to *M*.¹²

The last point is meant to illustrate the respect and manner in which natural laws "support" counterfactuals. Laws, being relationships between properties and magnitudes, go *beyond* the sets of things in *this* world that exemplify these properties and have these magnitudes. Laws tell us that quality *F* is linked to quality *G* in a certain way; hence, if object *O* (which has neither property) were to acquire property *F*, it would also acquire *G* in virtue of this connection between *F*-ness and *G*-ness. A statement of law asserts something that allows us to entertain the prospect of alterations in the extension of the predicate expressions contained in the statement. Since they make no reference to the extensions of their constituent terms (where the extensions are understood to be the things that are *F* and *G* in this world), we can hypothetically alter these extensions in the antecedent of our counterfactual ("if this were an *F* . . .") and use the connection asserted in the law to reach the consequent (" . . . it would be *G*"). Statements of law, by talking about the relevant properties rather than the sets of things that have these properties, have a far wider scope than any true generalization about the actual world. Their scope extends to those possible worlds in which the extensions of our terms differ but the connections between properties remains invariant. This is a power that no universal generalization of the form $(x)(Fx \supset Gx)$ has; this statement says something about the actual *F*'s and *G*'s in *this* world. It says absolutely nothing about those possible worlds in which there are *additional F's* or

different F's. For this reason it cannot imply a counterfactual. To do this we must ascend to a level of discourse in which what we talk about, and what we say about what we talk about, remains the *same* through alterations in extension. This can only be achieved through an ontological ascent of the type reflected in (6).

We come, finally, to the notion of explanation and confirmation. I shall have relatively little to say about these ideas, not because I think that the present conception of laws is particularly weak in this regard, but because its very real strengths have already been made evident. Laws figure in the explanation of their instances because they are not merely summaries of these instances. I can explain why this *F* is *G* by describing the relationship that exists between the properties in question. I can explain why the current increased upon an increase in the voltage by appealing to the relationship that exists between the flow of charge (current intensity) and the voltage (notice the definite articles). The period of a pendulum decreases when you shorten the length of the bob, not because all pendulums do that, but because the period and the length are related in the fashion $T = 2\pi\sqrt{L/g}$. The principles of thermodynamics tell us about the relationships that exist between such quantities as energy, entropy, temperature and pressure, and it is for this reason that we can use these principles to explain the increase in temperature of a rapidly compressed gas, explain why perpetual motion machines cannot be built, and why balloons do not spontaneously collapse without a puncture.

Furthermore, if we take seriously the connection between explanation and confirmation, take seriously the idea that to confirm a hypothesis is to bring forward data for which the hypothesis is the best (or one of the better) competing explanations, then we arrive at the mildly paradoxical result that laws can be confirmed *because* they are more than generalizations of that data. Recall, we began this essay by saying that if a statement of law asserted anything more than is asserted by a universally true statement of the form $(x)(Fx \supset Gx)$, then it asserted something that was beyond our epistemological grasp. The conclusion we have reached is that *unless* a statement of law goes beyond what is asserted by such universal truths, unless it asserts something that cannot be completely verified (even with a complete enumeration of its instances), it cannot be confirmed and used for predictive purposes. It cannot be confirmed because it cannot explain; and its inability to explain is a symptom of the fact that there is not enough "distance" between it and the facts it is called upon to explain. To get this distance we require an ontological ascent.

I expect to hear charges of Platonism. They would be premature. I have not argued that there are universal properties. I have been concerned to establish something weaker, something conditional in nature: *viz.*, universal properties exist, and there exists a definite relationship between these universal properties, *if* there are any laws of nature. If one prefers desert landscapes, prefers to keep one's ontology respectably nominalistic, I can

and do sympathize. I would merely point out that in such barren terrain there are no laws, nor is there anything that can be dressed up to look like a law. These are inflationary times, and the cost of nominalism has just gone up.^{13*}

■ | Notes

1. This is the position taken by Hempel and Oppenheim ([10]).
2. When the statement is of nonlimited scope it is already beyond our epistemological grasp in the sense that we cannot *conclusively* verify it with the (necessarily) finite set of observations to which traditional theories of confirmation restrict themselves. When I say (in the text) that the statement is "beyond our epistemological grasp" I have something more serious in mind than this rather trivial limitation.
3. Most prominently, William Kneale in [12] and [13].
4. I eliminate quotes when their absence will cause no confusion. I will also, sometimes, speak of laws and statements of law indifferently. I think, however, that it is a serious mistake to conflate these two notions. Laws are what is expressed by true lawlike statements (see [1], p. 2, for a discussion of the possible senses of "law" in this regard). I will return to this point later.
5. Popper ([17]) vaguely perceives, but fails to appreciate the significance of, the same (or a similar) point. He distinguishes between the structure of terms in laws and universal generalizations, referring to their occurrence in laws as "intensional" and their occurrence in universal generalizations as "extensional." Popper fails to develop this insight, however, and continues to equate laws with a certain class of universal truths.
6. Nelson Goodman gives a succinct statement of the functionalist position: "As a first approximation then, we might say that a law is a true sentence used for making predictions. That laws are used predictively is of course a simple truism, and I am not proposing it as a novelty. I want only to emphasize the Humean idea that rather than a sentence being used for prediction because it is a law, it is called a law because it is used for prediction; and that rather than the law being used for prediction because it describes a causal connection, the meaning of the causal connection is to be interpreted in terms of predictively used laws" ([7], p. 20–21). Among functionalists of this sort I would include Ayer ([2]), Nagel ([16]), Popper ([17]), Mackie ([14]), Bromberger ([6]), Braithwaite ([3]), Hempel ([10], [11]) and many others. Achinstein is harder to classify. He says that laws express regularities that can be cited in providing analyses and explanations ([1],

* This essay was written during the 1970s, a decade of high inflation. Nominalists deny that universals have any real existence, insisting that general terms such as *red*, *giraffe*, and *electrically charged* do not refer to universal properties, abstract objects, or Platonic forms. Typically, nominalists view the meaning of general terms as deriving from particular resemblances between particular things.

p. 9), but he has a rather broad idea of regularities: "regularities might also be attributed to properties" ([1], pages 19, 22)

7. I attach no special significance to the connective " \rightarrow ." I use it here merely as a dummy connective or relation. The kind of connection asserted to exist between the universals in question will depend on the particular law in question, and it will vary depending on whether the law involves quantitative or merely qualitative expressions. For example, Ohm's Law asserts for a certain class of situations a constant ratio (R) between the magnitudes E (potential difference) and I (current intensity), a fact that we use the " $=$ " sign to represent: $E/I = R$. In the case of simple qualitative laws (though I doubt whether there are many genuine laws of this sort) the connective " \rightarrow " merely expresses a link or connection between the respective qualities and may be read as "yields." If it is a law that all men are mortal, then humanity yields mortality (humanity \rightarrow mortality). Incidentally, I am not denying that we can, and do, express laws as simply "All F 's are G " (sometimes this is the only convenient way to express them). All I am suggesting is that when lawlike statements are presented in this form it may not be clear what is being asserted: a law or a universal generalization. When the context makes it clear that a relation of law is being described, we can (without ambiguity) express it as "All F 's are G " for it is then understood in the manner of (6).

8. On the basis of an argument concerned with the restrictions on predicate expressions that may appear in laws, Hempel reaches a similar conclusion but he interprets it differently. "Epitomizing these observations we might say that a lawlike sentence of universal nonprobabilistic character is not about these classes or extensions *under certain descriptions*" ([11], p. 128). I guess I do not know what being *about* something *under a description* means unless it amounts to being about the property or feature expressed by that description. I return to this point later.

9. Molnar ([15]) has an excellent brief critique of attempts to analyze a law by using epistemic conditions of the kind being discussed.

10. Brody argues that a qualitative confirmation function need not require that any E that raises the degree of confirmation of H thereby (qualitatively) confirms H . We need only require (perhaps this is also too much) that if E does qualitatively confirm H , then E raises the degree of confirmation of H . His arguments take their point of departure from Carnap's examples against the special consequence and converse consequence condition ([4], pages 414–418). However this may be, I think it fair to say that most writers on confirmation theory take a *confirmatory* piece of evidence to be a piece of evidence that *raises* the probability of the hypothesis for which it is confirmatory. How well it must be confirmed to be acceptable is another matter of course.

11. If the hypothesis is of nonlimited scope, then its scope is not known to be finite. Hence, we cannot know whether we are getting a numerical increase in the ratio: examined favorable cases/total number of cases. If an increase in the probability of a hypothesis is equated with a (known) increase in this ratio, then we cannot raise the probability of a hypothesis of nonlimited scope in the simple-minded way described for hypotheses of (known) finite scope.

12. If the law was interpreted as a universal imperative of the form described, the most that it would permit us to infer about Sally would be a counterfactual: If

Sally were one of the Presidents (i.e. identical with either Ford, Nixon, Johnson, . . .), then she would (at the appropriate time) have to consult Congress on matters pertaining to M.

13. For their helpful comments my thanks to colleagues at Wisconsin and a number of other universities where I read earlier versions of this paper. I wish, especially, to thank Zane Parks, Robert Causey, Martin Perlmutter, Norman Gillespie, and Richard Aquilla for their critical suggestions, but they should not be blamed for the way I garbled them.

■ | References

- [1] Achinstein, P. *Law and Explanation*. Oxford: Clarendon Press, 1971.
- [2] Ayer, A. J. "What Is a Law of Nature." In [5], pages 39–54.
- [3] Braithwaite, R. B. *Scientific Explanation*. Cambridge, England: Cambridge University Press, 1957.
- [4] Brody, B. A. "Confirmation and Explanation." *Journal of Philosophy* 65 (1968): 282–299, Reprinted in [5], pages 410–426.
- [5] Brody, B. A. *Readings in the Philosophy of Science*. Englewood Cliffs, N.J.: Prentice Hall, 1970.
- [6] Bromberger, S. "Why-Questions." In [5], pages 66–87.
- [7] Goodman, N. *Fact, Fiction and Forecast*. London: The Athlone Press, 1954.
- [8] Harman, G. "The Inference to the Best Explanation." *Philosophical Review* 74 (1965): 88–95.
- [9] Harman, G. "Knowledge, Inference and Explanation." *Philosophical Quarterly* 18 (1968): 164–173.
- [10] Hempel, C. G., and Oppenheim, P. "Studies in the Logic of Explanation." In [5], pages 8–27.
- [11] Hempel, C. G. "Maximal Specificity and Lawlikeness in Probabilistic Explanations." *Philosophy of Science* 35 (1968): 116–133.
- [12] Kneale, W. "Natural Laws and Contrary-to-Fact Conditionals." *Analysis* 10 (1950): 121–125.
- [13] Kneale, W. *Probability and Induction*, Oxford: Oxford University Press, 1949.
- [14] Mackie, J. L. "Counterfactuals and Causal Laws." In *Analytical Philosophy*. (First Series). Edited by R. J. Butler. Oxford: Basil Blackwell, 1966.
- [15] Molnar, G. "Kneale's Argument Revisited." *Philosophical Review* 78 (1969): 79–89.
- [16] Nagel, E. *The Structure of Science*. New York: Harcourt Brace, 1961.
- [17] Popper, K. "A Note on Natural Laws and So-Called 'Contrary-to-Fact Conditionals.'" *Mind* 58 (1949): 62–66.

Necessities and Universals in Natural Laws

1 | Prologue

How do laws of nature differ from cosmic coincidences? This is a question very familiar to philosophers of science, and answers of two sorts still vie for their allegiance. One sort locates the difference in what laws say, the other "in the different roles which they play in our thinking", as Braithwaite's *Scientific Explanation* put it (1953: 295). In Chapter 9 of that book, Braithwaite developed and defended a classic answer of the second sort: the difference, he says there, lies in why we believe laws, not in what they say. In the quarter century since then, other answers of the same sort have been devised: Hesse presents one in [Hesse (1980)]. But since then also, answers of the first sort have again come into fashion. The revived fashion has mostly been for reading laws as saying how things must be; but some, more recently, have read them instead as relating not things but properties of things to each other. Hesse notes these fashions and rejects them, to my mind rightly, but she does not elaborate her reasons. It seems to me therefore that I can best complement her article by inspecting these fashions' argumentative cut, to see if they do indeed fit better than her and Braithwaite's Humean gear. Only first I shall build the problem up in my own way, to provide a lay figure to hang the garments on.

2 | The Problem

Certified laws of nature are the primary products of scientific thought and observation. They embody the generalized knowledge which science yields; they supply explanations and predictions of events; and they un-

FROM D. H. Mellor, ed., *Science, Belief and Behaviour: Essays in Honour of R. B. Braithwaite* (Cambridge: Cambridge University Press, 1980), 105–25.

derlie the design of most modern artefacts. To take just three obvious examples: our human life has been much altered in this century by the discovery and applications of laws governing plant genetics, aerodynamics and electromagnetic radiation.

Laws differ widely in their subject matter, importance and complexity. What they have in common is generality. A law says that *all* things or events of some kind have a certain property or are related in a certain way to something else. If the law is statistical, the property is having a chance of having some other property or of being related to something else. It is, for example, a law that all light has the property of going at the same speed in a vacuum; and it is a statistical law that all atoms of the most common isotope of radium have the same chance (fifty fifty) of turning into something else within their half life of 1622 years.

What needs certifying about a law is its truth. We cannot know that all light goes at the same speed in a vacuum unless it truly does so. Its constant speed will not serve to explain or predict anything if its speed is not in fact constant. And it is unsafe to base the design of artefacts on what is not the case. We know of course that even a certified law may turn out to be false. But without good reason to think it true, we lack good reason to employ it as we do. This is why we do not call something a law unless we think it true, so that a false generalization cannot be a law, although it may be "lawlike": *i.e.* such that it would be a law if only it were true.

Certifying the truth of some laws presents no problem. These are the analytic laws, those whose truth follows from the meanings of the terms they are couched in. There are more reasons than one for laws being analytic. A law may be analytic because it is used to define one of its terms. Newton's laws of motion, for example, may well be analytic because between them they define the Newtonian concepts of force and mass. Or a law, not originally analytic, may become so successful and theoretically important that its terms change their meaning to make it analytic. For this reason it is now arguably analytic that light is electromagnetic radiation, although that could not have been the case when the electromagnetic theory was first conjectured to apply to light. Then, we could easily have envisaged observing light to go faster or slower, for example, than the theory can be shown (by measuring the ratio of electromagnetic to electrostatic units) to require electromagnetic radiation to go. Nowadays we should take such an observation to show some error in the theory rather than question the law that light is electromagnetic radiation.

But even if some laws are analytic, most laws are not, and these are the ones that concern me. It does not follow from the meanings of the terms involved that radium's half life is 1622 years, nor that benzene is as insoluble in water as it is. Nothing semantic prevents a little more benzene sometimes dissolving in water, or some piece of radium having a rather different half life. How then can we certify the truth of what the law says,

namely that these things never happen? We cannot see that they never do, if only because at no time can we see that they never will do in the future. We cannot directly perceive the truth of nonanalytic laws. At most, our senses can show us some of a law's past instances, and then only instances of laws about relatively observable properties of things and events. We can observe the speed of this or that ray of light and, perhaps indirectly, the half life of this or that piece of radium; but not all the things and events, past, present and to come, to which the law applies.

The problem then arises why a supposed law should be expected to hold in instances as yet unobserved; in short, Hume's problem of induction. Unlike Popper and his followers, I believe that induction does present a genuine and serious problem, which needs solution and has not yet been solved; although I believe Braithwaite's (1953: Ch. 8) attempted solution is along the right lines.* But wherever its solution lies, Hume's problem does not arise only incidentally for laws of nature. On the contrary, it is an inevitable concomitant to their role in supplying predictions. To make a prediction is to anticipate, rightly or wrongly, the result of making an observation; to say or just to expect, for example, that a bomb will explode before we see it do so. Whatever purports, as a law does, to justify such an expectation necessarily arouses Hume's problem. Only a generalization certified by observing all its instances would be free of inductive pretensions, and such a generalization is not much use for predicting things. It might indeed have some use: one might accept it on someone else's authority and use it to predict some instances one had not observed oneself. But real laws are used amongst other things to predict the results of future observations, and these are not yet available to anyone to certify the law with (see Mellor 1979). Real laws therefore undeniably need inductive support.

The other philosophical problem which laws of nature present is the one that concerns us. It is less obvious than the problem of induction, but perhaps more tractable: what exactly do laws say? I have taken them to be generalizations, and there is not much doubt of that. The debatable question is whether laws are more than generalizations, and if so, what more. Now if giving laws one content rather than another made the problem of induction soluble for them, this would be a strong argument for giving them that content. But since I believe no such solution is presently available for any credible content, I must look to other arguments. Hume's problem does, however, provide a reason for preferring weak readings of

* The problem of induction is discussed in the readings by Lipton and Popper in chapter 4 and in the accompanying commentary. For Mellor's more recent views on induction, see D. H. Mellor, "The Warrant of Induction," in *Matters of Metaphysics* (Cambridge: University of Cambridge Press, 1991), 254–68.

natural laws. The less a law says, the less there is to be certified in claiming it to be true.

The weakest reading seems to be the obvious one I have already given:

1 All *F*s are *G*s,

where *F* and *G* are properties of things or events. They may be relational, comparative or quantitative properties; in statistical laws *G* will be some determinate chance of having another property. (1) is of course a very simple form of law, but it will do; it has all the relevantly problematic features of more complex forms. But before discussing its supposed deficiencies, some preliminary points need to be made clear.

First, as my examples have already illustrated, the 'are' in (1) is to be taken tenselessly. The law applies to all *F* items in the universe, past and future as well as present. The laws of radioactivity do not just give radium's present half life; they say what it always was and always will be. Now some of what we take to be physical constants, such as the half life of radioelements, might indeed turn out to depend on the age of the universe. But then the true laws of radioactivity would say what the dependence was. Those laws would, like all other true laws, apply at all times; the values of our supposed constants at particular epochs being merely special cases of the general laws.

Secondly, I take it that anything in the universe is definitely either *F* or not *F*, either *G* or not *G*. This is not an uncontentious claim. Some have been led to deny it of so-called "vague" properties like being bald, because of its seemingly absurd consequences (for example, that at some point adding just one hair to a bald man's head gets rid of his baldness). Others have been led to deny it of some things and events in the future, either because they want the future to be open, at least in some respects, to being made definite by human decision and consequent human action or because of problems raised by quantum mechanics. They think it cannot now be the case, for example, that I shall definitely either be dead or be alive next year, if it is still open to me and others to settle the matter by what we decide to do between now and then. I think that these are both inadequate grounds for denying that everything is definitely *F* or not *F*, but I shall not argue the point here. (On the first, see Cargile 1969; on the second see Mellor 1981. I also think my being wrong in either case would make little difference to the ensuing discussion, but I shall not argue that either.)

Thirdly, I exclude from the range of *F* and *G* factitious properties such as Goodman's (1965) notorious "grue" (= green if the item is inspected before a specified time, otherwise = blue). I hope and believe criteria can be given to rule out these phoney properties (see for example Hesse 1974: Ch. 3); but in any event all parties agree that they are phoney, and I shall take their exclusion for granted.

I should however emphasize that I do not mean to restrict F and G to physical, as opposed to psychological or social, properties. Some philosophers (e.g. Davidson 1970; McGinn 1978) deny the existence of laws relating nonphysical properties; but largely because they mistake laws to involve necessities of the kind I shall be concerned to dispute and which they correctly perceive to be absent from mental and social generalizations. Anyway the point should be left open here; so if I stick to physical examples, it is only to avoid irrelevant controversy, and not because I think there are no others.

With this preamble, we may now ask what, if anything, is wrong with (1) as a reading of laws of nature. To see what seems to be wrong, we must look at (1)'s consequences in special cases, particularly the case, on which Braithwaite concentrates, where nothing in the world is F .

One might imagine that it did not matter what follows from (1) when nothing is F , but it does. Let us call a law 'vacuous' in that case. Many important laws are vacuous in this sense. The most famous one is Newton's first law of motion, that bodies acted on by no forces are at rest or move at a constant speed in a straight line. The law is central to Newtonian mechanics, but Newton's own gravitational theory implies its vacuity, since the theory says that all bodies exert gravitational forces on each other. No doubt Newton's laws of motion are peculiar, since as already remarked they may well be analytic. But Newton's first law illustrates a vacuity which is shared by many laws that are in no way analytic. There is in particular a multitude of nonanalytic laws quantifying over determinate values of continuously variable determinables: for example, the laws relating the vapour pressure of substances to their temperature. Each determinate value of these determinables yields another law as a special case, such as the law giving the boiling point of water at atmospheric pressure. Now there are infinitely many different temperatures and pressures, and hence infinitely many of these derived laws, all with mutually incompatible antecedents (nothing can be wholly at two different temperatures or pressures at the same time). Although the temperature and pressure of any given mass of water will vary continuously with time, there are many temperatures which no mass of water ever reaches: temperatures, for example, so high that water would decompose before it reached them. At any rate, so far as these derived laws are concerned, it is entirely accidental whether any water ever is at the temperatures and pressures they apply to. Consequently they must certainly be so construed as to make equal sense whether they happen to be vacuous or not (cf. Ayer 1956: 224–5 [818]).

In particular, it seems obvious that mere vacuity should not settle the truth of a law regardless of its content. But a lack of F s makes 'All F s are A s' true for any A , including both $A = G$ and $A = \text{not-}G$. If there never is any water at some temperature T , statements crediting all water at that temperature with any pressure whatever all come out true. That seems

absurd; so vacuous laws should be read as saying something other than 'All *F*s are *G*s'. The question is what.

The obvious answer is that a vacuous law says

2 If there were *F*s, they would be *G*s.

But there are objections to (2). One is that it appears to imply that there are no *F*s, whereas laws, even if they happen to be vacuous, certainly do not claim to be. We could in reply say that (2) is not to be read as having this implication; and this stipulation can indeed be given some independent rationale. A case can be made for saying that the implication is not part of what (2) says, but follows rather from applying general rules of discourse: namely, not to mislead, and to be as informative as possible (see Mackie 1973: 75–7). These rules dictate that one should not say (1) when the law is known to be vacuous, since (1) is no more true than is any other generalization starting 'All *F*s are . . .'. To pick out as a law the generalization which relates *F* especially to *G* in these circumstances, one must have some reason other than its truth. The reason may not be specified, but the fact that there is one is signalled by using (2) instead of (1). Consequently, even if the law says no more than (1), (2) would normally be used when, but only when the law is known to be vacuous. So (2) will indeed signal its user's knowledge of the law's vacuity, even though that is no part of what (2) is being used to say.

This is one of the arguments which can be used to defend Humean accounts of laws as saying no more than (1). It still leaves the problems of saying what reason there is to link *F* and *G* as a law does when there are no *F*s, and why (2) should be the right way to signal this reason. These are among the problems that have exercised Braithwaite and his Humean successors. But since my concern here is with their rivals, I shall concentrate instead on recent attempts to solve the problem of vacuous laws by giving (2) some assertible content over and above (1).

Laws, I have remarked, do not claim to be vacuous, even if they are; and ideally, they should say the same thing whether they are vacuous or not. It will hardly do to make laws say (2) if they are vacuous and (1) if they are not. A law cannot say (1) in both cases, we are supposing; can it say (2)? We have dealt with the obvious objection by removing (2)'s counterfactual implication (that there are no *F*s), which would have made all nonvacuous lawlike generalizations false regardless of their content. What can be said positively in favour of the suggestion?

Consider the universe of non-*F* things or events of which a vacuous law says that if they were *F*s they would be *G*s. It is surely immaterial to this supposed fact about these things or events that there happens to be nothing else which is *F*. So perhaps we should take the nonvacuous law also to say of every non-*F* thing or event that if it were *F* it would be *G*. But again, the law itself does not assert that these things or events are not

F. It should say the same of all things or events, whether they are *F* or not. Let us therefore take a law to say of every thing or event *x* that

3 If *x* were *F* it would be *G*.

(Those who believe in possible as well as actual things and events may take '*x*' to range over them too.) I shall take the problem for our non-Humeans to be that of saying what (3) means in this case.

I shall not demand of them a general analysis of so-called 'subjunctive' or 'counterfactual' conditionals like (3). A general analysis would of course have to cover those that we are supposing to give the content of natural laws. But I am not convinced that other uses of these conditionals are homogeneous enough with this one to shed much light on it. In most other uses, for example, (3) might very well imply that *x* is not *F*, which we have seen it cannot do here. Or again, to make (3) true of an *x*, it may often suffice for that *x* to be *F* and also *D*. Lewis's influential analysis, for example, takes this more or less for granted, and the way he reluctantly accommodates possible exceptions (1973: 29) will certainly not cope with natural laws. Yet natural laws must be exceptions: it might be a coincidence that an *x* is both *F* and *G*, and not a matter of natural law at all. So in this case it must take more than that to make (3) true of any *x*. And as our consideration of vacuous laws has shown, the extra cannot be that all other *F*s are *G*s too, for there might just as well be no other *F*s. So whether the law is vacuous or not, the truth of (1) will not suffice to make (3) true of everything. But what more than (1) can a law say?

3 | Possible Worlds

The traditional non-Humean answer is that natural laws are or express necessities of some kind: what makes (3) true of everything is that *F*s not merely are *G*s, they have to be; (1) is not merely true, it is necessarily so. Conceptions of law as what Kneale (1949) called 'principles of necessitation' are of course by no means new. The problem with them is to justify the idea of necessity they invoke and to show how it explains the universal truth of (3). Of late years, the development of so-called "possible world semantics" has made that problem look more tractable, and thus encouraged a revival of the idea that natural laws are necessary truths. It has done this by providing a systematic way of saying what makes statements of necessity (and of possibility) true. So in particular we might hope to find in it an acceptable way of saying what makes necessary natural laws true.

The basic concept of this semantics is that of a possible world. A possible world is a way the world might be, or might have been. There are many such ways, and therefore many possible worlds, of which the actual world is just one. Possible worlds are distinguished by what the facts

are supposed to be in them: if the supposed facts differ at all, so do the worlds. I might, for instance, die in various ways, and, for each way, at various ages. So there are many possible worlds in which I expire of, say, cirrhosis (or my counterpart in that world does so: see Lewis 1973: 39), and these differ amongst other things according to my or my counterpart's age at the time. In general, a statement which might be true, but fails to specify every detail of the universe, will be true in many possible worlds, differing amongst themselves in the details left unspecified.

Having in some such manner as this grasped the idea of possible worlds, and reified them, one can turn round and give, as the truth conditions of a statement, the set of possible worlds in which it is true. That is how possible world semantics offers to give the meaning of various kinds of modal statements, and in particular of statements of necessity and possibility. How enlightening this conceptual round trip is, from what might be the case, to what is the case in a possible world, and back again, is a very moot point, but one that can be waived while we see how well the concept copes with the supposed necessity of natural laws.

It follows at once from the definition of a possible world that a statement which might be true is one that is true in some possible world. Hence statements which have to be true are those which are true in all possible worlds. In particular, for (1) to be necessarily true is for it to hold in all the worlds there might be or might have been. Is that really what a natural law claims?

Suppose it is: does that solve the problem of vacuous laws and explain (3)'s being true of everything in the actual world? Consider again the case where there are no actual *F*s. The law does not say there are none, and it is tempting to suppose there always might have been. If that were so, then, on this view of laws, (1) would have to be true not only in this world, but also in worlds containing *F*s where its truth would not be the trivial consequence of vacuity it is here. And that would certainly distinguish (1) as a law from other vacuously true generalizations.

But this account depends on the possibility of there being *F*s; and, on this view of laws, there will often be no such possibility. I have cited the example of high temperature instances of the vapour pressure law for water that are vacuous because water decomposes before it reaches those temperatures. Now, that water decomposes below these temperatures is itself a natural law and so, on this view of them, necessary. Consequently these high temperature instances of the vapour pressure law not only are vacuous, they have to be. There could be no water at such temperatures. But that is to say there are no possible worlds in which these instances are not vacuous; and therefore none in which the truth of this instance of (1) is other than a trivial consequence of vacuity.

So the idea of laws being true in all possible worlds does not solve the problem of vacuous laws. Nor, for much the same reason, does it explain why (3) is true of everything in this world. Again, it would if

anything, a , in this world might have been F even if it isn't. Then there would be possible worlds in which a (or some counterpart of a) is in fact F ; and in all these worlds it, like every other F , is G . Where that is so, it seems to me undeniable that (3) is true of a . However, for any F there will be many a s of which it is quite incredible that they might have been F . Take the law that in a vacuum all light goes at a constant speed—which is to say that all photons do. It is true then, of anything at all, that it would go at that speed in a vacuum if it were a photon. But this is not to say of everything that it might have been a photon. There is no possible world in which I am (or any counterpart of me is) a photon; and *a fortiori* none in which, as a photon, I (or any counterparts of me) travel at the speed of light. That is not, I believe, what makes this instance of (3) true of me. Yet I believe it is true of me, since I believe the law; and there is surely no inconsistency in my combining these beliefs.

For subjunctive conditionals like (3) to be true, their antecedents do not have to be possible. This is blatantly obvious in *reductio ad absurdum* proofs, where the truth of a subjunctive conditional is actually used to prove that its antecedent is *not* possible. One and one cannot make three precisely because, if they were to, something impossible would be the case. It should be almost as obvious that conditionals which give the content of natural laws likewise do not imply the possibility of their antecedents being true. The vapour pressure example shows at least that they cannot both do this and themselves be necessary truths. And I have given elsewhere (1974: 173) the example of safety precautions at a nuclear power station, which are supposed to make impossible the conditions under which, as a matter of natural law, the fuel would explode. It is ridiculous to maintain that the success of these precautions would disprove the very law that makes them necessary.

I am not sure why (3) should be so often thought to imply that x might be F . The reason may well be the same for taking (3) to imply that x is not F : namely, that it is customary to reserve subjunctive conditionals for use when their antecedents are believed to be false but possible. We see, however, that this custom is not invariable, and have in any case seen reason (see p. 851) not to make such a custom part of a conditional's meaning. So however natural the thought may be, it is mistaken, at least of the conditionals implied by natural laws. But the mistake is very widespread and of long standing, and it has had serious consequences. It has bedevilled the analysis of disposition statements, as I argued in §9 of my (1974). It has likewise afflicted discussions of free will, in which 'I could have done X ' is frequently equated with 'I would (or could) have done X had I chosen to'. But it obviously does not follow from the latter that I could have done X , since it obviously does not follow that I could have chosen to.

The common confounding of conditional statements with statements of possibility has thus had ill effects in more than one area of philosophy.

The ill effect here has been that possible world semantics have been mistakenly thought to give sense to the idea that natural laws are, or assert, some kind of necessity.

4 | Natural Necessity

Laws might however still be necessary even if possible world semantics fails to say what makes them so. What makes (3) true of everything might still be that nothing could be both *F* and not *G*, whether or not it could be *F*. But it is not at all obvious that this is so. Subjunctive conditionals are not in general made true by necessities. Suppose that if I were to go to London I would go by train. This does not mean that I could not go any other way, merely that I would not. Lewis's (1973) treatment of subjunctive conditionals recognizes this fact about them: the consequent does not have to be true in all the possible worlds the antecedent is true in, only in those most like the real world.

Still, I have insisted that conditionals like (3) which follow from laws are a special case. In particular, it does not suffice for their truth that their antecedents and consequents are true; whereas my going to London by train may well make it true that, were I to go, I would go that way. So perhaps (3) does need some necessity to make it true of everything, even if conditionals in general do not.

But most natural laws seem to be contingent. Apart from those that are definitions, and those whose success has made them analytic, any law might have been false. We could have come across a counterexample to it, and we still could, even if we never will or would. That seems at any rate to be why we need to test our supposed laws by observation: things could be other than the law says, so we need to look and see whether or not they are. I believe, for example, that light could have gone in a vacuum at other than its constant speed, even if no photon ever does and even if nothing, were it a photon, ever would. So on the face of it, conditionals like (3) no more exhibit necessity than does the conditional about my going to London by train.

Attempts have been made to explain away the apparent contingency of natural laws. One attempt, which need not detain us long, distinguishes logical necessity and possibility from their natural or physical counterparts. It is logically possible for *F*s not to be *G*, but not naturally or physically possible. But all 'physically possible' means is 'consistent with natural law'. So to say that something is physically necessary is merely to say that some law entails it. Whether the law says it has to happen, and whether the law itself has to be true, remain entirely open questions.

A more serious attempt distinguishes between metaphysical and epistemic necessities (Kripke 1971: 150–1; Dummett 1973: 121); that is, be-

tween being necessary and being knowable *a priori*. Laws appear to be contingent because they cannot be known *a priori*. They cannot be proved in the way the truths of logic and mathematics can. We need to look and see what the world's laws are, and it may always turn out that what was thought to be a law really is not one. The *F*s we have seen to be *G* may mislead us into believing they all are, even though some future ones are not. It is consistent with all we have seen that there should be *F*s which are not *G*. That is the epistemic possibility of a supposed law being false; and something like it exists in mathematics. There too, special cases may mislead us into believing a mathematical generalization to which there are in fact counterexamples. Now, recognizing this possibility in mathematics does not diminish our belief in the necessity of mathematical truths: if the generalization is true, it could not have been otherwise. It is likewise conceivable that natural laws, if true, are necessarily so, even though we may be mistaken in what we suppose the true laws to be.

The apparent contingency of natural laws could undoubtedly be explained away like this if there were good reason to think them necessary: but is there? The analogy with mathematics certainly does not give one. If Goldbach's conjecture proves true, any attempt to suppose it false will eventually lead to contradiction (that of course being one way of proving it). In that case no consistent description could be given of a world in which the conjecture was true. That is, there is no such possible world. We might therefore explain the conjecture's necessity, if true, as truth in all possible, *i.e.* coherently conceivable worlds; since conceivability is a notion arguably more basic than necessity and intelligible independently of it. But no such case can be made for the corresponding conception of natural necessity. As Hume insisted, there is no difficulty in conceiving a natural law to be false: since it is not analytic, no contradiction ensues. A perfectly coherent description can be given of a world containing *F*s that are not *G*. The only ground for thinking such a world impossible would be that the law which would be false in it is not only true but necessary; and this is the very fact that needs to be established and explained.

5 | Essences

Arguments have recently appeared for the metaphysical necessity of some laws, namely those specifying essential properties of natural kinds. An essential property of a kind is one which nothing of that kind can lack. So if being *G* is of the essence of a kind *F*, the law that all *F*s are *G*s will be a necessary truth. The exemplars most widely touted by advocates of essences concern the microstructure of kinds: the atomic number of gold, the molecular constitution of water, the genetic makeup of plant and animal species, and the mean kinetic energy of gas particles at a given tem-

perature. The question is: why suppose that these, or any other, properties of kinds are essential?

Two sorts of arguments have lately been adduced for essences, and hence for the necessity of the corresponding laws. Both employ possible world semantics; neither therefore proves more than that some generalizations hold in all possible worlds, and we have seen in [section] 3 that this is not enough to serve our turn. But the arguments repay scrutiny nonetheless, since there is more to them than the possible world jargon they are couched in.

One argument, due to Putnam (1975), infers essences from a mechanism for fixing what things or events a kind predicate ('*F*') applies to, *i.e.* its extension. This mechanism fixes what things are, or might be, *F*s in two stages. First, there are archetypal actual *F*s (*e.g.* paradigm specimens of gold or water): things that have to be *F* if anything is. Second, to be *F*, anything else has to have a suitable 'same-kind' relation to these archetypes. What this means is that it has to share some property with them — apart of course from the property *F*. What the same-kind relation is, for any category of kinds, it is for empirically testable scientific theories to say. The relations are not discoverable *a priori*, and in particular they do not follow from the meanings of the predicates involved: the laws giving the essences of kinds are not supposed to be analytic. But any shared property *G* which a same-kind relation picks out will be an essential property of the kind since, Putnam assumes (1975: 232), the relation is an equivalence relation holding across all possible worlds. Thus not only are actual *F*s all *G*s, all possible *F*s are: so (1) in this instance is true in all possible worlds.

I have elaborated elsewhere (1977) my reasons for rejecting this argument. Briefly, the extensions of real natural kinds do not in the first place depend on archetypes in the way Putnam's mechanism requires. And, in the second place, even if they did, his mechanism would still not produce essences. To produce an essence, the same-kind relation must be transitive, in order to ensure that all possible *F*s share the *same* property *G* with each other. But the mechanism does not need a transitive relation, since what makes things *F*s in other possible worlds is their sharing some property other than *F* with the archetypal *F*s in this one, and there is nothing to say this shared property must be the same in every possible world. For Putnam to claim the same-kind relation to be transitive, which he does in taking it to be an equivalence relation, is for him gratuitously to assume the essentialist conclusion he is out to prove. His mechanism in fact gives us no reason to think any instance of (1) true in all possible worlds. And since in any case only those giving essences are in question, Putnam's theory, even if it worked, would not solve the general problem of distinguishing laws from universal coincidences.

The same of course is true of Kripke's (1971, 1972) argument for essences; but that too we must look at, since solving our problem for some laws would at least be better than solving it for none. Kripke's argument

is quite different from Putnam's. Kripke takes laws giving essences to be identity statements: the law 'Water is H_2O ' he takes to say not merely that anything, were it water, would be H_2O , but that being water and being H_2O are one and the same property. But identity is a necessary relation, in the sense that nothing could fail to be identical to itself. So being water and being H_2O are the same property in all possible worlds, not only in this one. Nothing that could be water could fail, were it water, to be H_2O .

This of course is the merest sketch of Kripke's argument. He has, for example, also to show that 'water' and ' H_2O ' are what he calls 'rigid designators', i.e. that they refer to the same stuff in any possible world it exists in. Otherwise the identity statement, since it is not analytic, might be true without being necessary, as for example 'Water is the most powerful solvent' is. ('The most powerful solvent' is not a rigid designator: it refers to whatever the most powerful solvent is, which in a world restricted largely to oil products would not be water.) As I have abbreviated Kripke's argument, so I shall abbreviate the objections raised in my (1977). The chief objection is that the argument, like Putnam's, blatantly begs the question: for being water and being H_2O to be the same property at all, never mind necessarily, the predicates ' \dots is water' and ' \dots is H_2O ' must already be coextensive in all possible worlds. This is not a conclusion to be derived from the necessity of the identity: it is built into the identity as a premise. Granted, 'Water is H_2O ' states a true law, and it has the form of an identity statement. But it is clearly only a variant of 'All water is H_2O ', which does not have that form. At any rate, the identity of these properties only follows if 'water' and ' H_2O ' are rigid designators, i.e. could not refer to different properties. But to believe this, one needs already to believe what the argument from this premiss is supposed to show: namely, that there could not have been some samples of water of a different molecular constitution.

Kripke, like Putnam, fails to establish the existence of essences. The microstructural exemplars which give their doctrine its spurious appeal indeed have a special status in science, but not the status of essences. They are special because they are central to our current scientific theories; but that, I have argued elsewhere (1977: §7), is quite a different matter from being necessary features of the world.

6 | Universals

The properties of being water and being H_2O do not stand in the necessary relation of identity. Perhaps, however, as Armstrong (1978a: Ch. 24), Dretske (1977) and Tooley (1977) have suggested, these universals stand in some contingent relation which makes it a law that all water is H_2O . This relation, that is to hold between the properties F and G whenever 'All F are G ' is a law, Armstrong and Tooley call 'nomic necessitation'; I shall

call it 'N'. *F* and *G* have to be differently related if the law is that no *F* are *G* or that all *F* have a chance *p* of being *G*. But *N* will do for now: if it works, the other relations will; if not, nor will they.

This suggestion requires a realist view of at least those universals which are related by natural laws: for *N* to relate *F* and *G*, these properties must exist. That of course is debatable, but suppose for the moment it is true. Then *FNG* is by definition the fact that makes 'All *F*s are *G*s' a law. This is a contingent fact, and not only because *F* and *G* might not exist. *F* and *G* could quite well exist without 'All *F*s are *G*s' being a law: laws do not relate every property to every other property. Being water and being at 100°C, for example, are properties that enter into laws, yet no law relates them to each other. But though it is not, it might have been a law that all water is at 100°C; just as it might not have been a law that all water boils at that temperature at atmospheric pressure. Apart from analytic laws, therefore, it is quite contingent that *N* relates any particular *F* and *G*.

To do its job, *N* has not only to make *G*s out of actual *F*s, it has to make (3) true of everything, i.e. to be such that anything, were it *F*, would be *G*. Since this is all it has to do and be, one might think that postulating *N* is more a relabelling of the problem than a solution to it. But that would not be a fair response. There is a dearth of candidates for making (3) universally true, as we take it to be. If *F*, *G* and *N* would between them make it true, that may well, as Tooley (1977: 262) urges, be reason enough to believe in them.

After all, we already invoke properties to make conditionals true. The inertial mass, *m* kilogrammes, of a thing *a* at time *t* makes true all the conditionals of the form

- 4 If *a* were subjected at *t* to a force of *f* newtons, it would then accelerate in the direction of the force at *f*/*m* metres/second².

Any of these conditionals is in fact a generalization about events, namely that, if they were subjectings of *a* to a (specific) force *f* at *t*, they would be (or be shortly followed by) accelerations of *a* of magnitude *f*/*m*. These generalizations are just like the conditionals (3) entailed by laws, except that they are restricted to the individual *a*. We think them true, and a fact is needed to make them so: and the requisite fact is that *a* has mass *m* at time *t*. This is all the property of inertial mass amounts to: a truth-maker, as Tooley puts it, for conditionals like (4); and I have argued elsewhere (1974) that all properties of things are just truth-makers for such conditionals. But if we believe in properties *F* and *G* because they are needed (and suffice) to make conditionals like (4) true, why jib at accepting *N* when it is likewise needed and (with *F* and *G*) suffices to make conditionals like (3) true?

Here, however, the crucial difference between (3) and (4) emerges: (4) entails that *a* exists. Without *a*, '*a*' would have no reference, and I do not see what (4) could then mean, nor how in particular it could be true.

So if (4) is true, a exists; so the fact that a has mass m is always available to make (4) true. But it is by no means so clear that F and G exist whenever (3) is true of everything. The law that all F are G , it is agreed, may be vacuous: and if it is, there are no F s. Now if, as many (including Armstrong) suppose, properties and other universals need instances, then without F s there will be no F . But without F there will be no fact FNG to make (3) true of everything; and the problem of accounting for vacuous laws will remain unsolved.

Perhaps then universals need no instances. Concepts certainly do not: we can have the concept of a unicorn without there being unicorns. But universals are not concepts: concepts, if anything, are parts of our thought or our language; whereas universals, if anything, are parts of the world whether or not it contains any thought or language or concepts. No doubt concepts are closely related to universals, but it is not safe to assume that universals can dispense with instances just because concepts can. That remains an open question.

Tooley takes it for granted that universals need no instances, since he uses a particular example of a vacuous law to argue by elimination "that it must be facts about universals that serve as the truth-makers for basic laws without positive instances" (1977: 672), going on to ask rhetorically: "if facts about universals constitute the truth-makers for some laws, why shouldn't they constitute the truth-makers for all laws?" Armstrong, by contrast, holds a "Principle of Instantiation: For each N -adic universal, U , there exist at least N particulars such that they U " (1978: 113); but he offers nonetheless to cope with Tooley's example of a vacuous law. Now if Armstrong really can supply enough universals to make vacuous laws true, without violating his principle of instantiation, we may not have to decide whether universals do in fact need instances; but can he?

In Tooley's example, as Armstrong puts it, just two out of several types of particle happen never to meet; so the law governing their interaction is vacuous. Nevertheless particles of other types meet, so that the universal, meeting (M), exists; as do these two mutually evasive particle types (A and J). Armstrong can claim therefore that the law, despite being vacuous, "holds in virtue of the universals [A , J , M] being what they are" (1978a: 157). This solution, however, is only available for special cases of vacuous laws. For a start, it only works here because particles of other types do meet, thereby ensuring the existence of the universal M . Now if the law governing A and J particle interactions does not ensure their meeting, it can hardly ensure the meeting of other types of particles; and if A and J particles can fail to meet, so can others. If no particles ever met, the laws of all their interactions would still be true, but the universal M would not, for Armstrong, exist to make them so.

More seriously, not only might there be no meetings, there might be no A or no J type particles. Yet the law could still be true and important, even if there were nothing it applied to. We have seen that to be the case

with Newton's first law of motion, and the vacuous instances of vapour pressure laws. For these, and indeed for the bulk of vacuous laws, Armstrong's principle of instantiation does deprive him of the universals he needs as truth-makers. The Tooley case he discusses happens to be of the only sort he can cope with, and it is worth drawing out what makes cases of this sort amenable to Armstrong's treatment: namely, that in them there exist things with properties which make certain generalizations true. These are in fact generalizations of conditionals like (4) above. Consider that for many determinate values of the determinable force f , (4) is vacuous: a can only be subjected to one (net) force at any one time, and there will be many forces a never experiences. Yet a 's always having mass m suffices to make all these vacuous generalizations true. And so it is with Armstrong's A and J particles. While they exist, they are disposed to interact as the law says, whether they ever actually meet or not. Armstrong's universals A and J are just conjunctions of such dispositions (see Mellor 1974), and can thus be truth-makers for those laws whose vacuity results merely from the dispositions of actual things failing to display themselves. But not for the more important cases in which laws are made vacuous by the non-existence of the things themselves; and hereafter I will reserve the term 'vacuous' for such cases.

Since vacuous laws will in fact defeat the Armstrong-Dretske-Tooley account if universals need instances, we have after all to consider whether they do. I follow Ramsey in taking particulars and universals to be simply parts of facts picked out in order to generalize. For example, "It is not ' aRb ' but ' $(x)xRb$ ' which makes Rb prominent. In writing ' $(x)xRb$ ' we use the expression ' Rb ' to collect together the set of propositions xRb which we want to assert to be true" (Ramsey 1925: 28–9). To recover a proposition from this set, we need to know what an instance of xRb is, i.e. we need criteria for identifying the items, such as a , which have been quantified over. But we do not need separate criteria to identify Rb . Given a , Rb is just the remainder of the fact aRb . If it were an independently identifiable constituent, then, as Ramsey says (1925: 23–4), $a(Rb)$ would differ from $(aR)b$ and $a(R)b$, because these facts would have different constituents; and this is absurd.

Similarly, if we form the doubly general ' $(x)(y)xRy$ ', we must regard the universal R as just the common part of all the facts thus collected: the fact aRb minus a and b . And since in forming the law that all F s are G s we at least collect whatever facts such as Fa there may be, the universal F must likewise just be the common part of all such facts. At least in laws, therefore, a universal must be regarded as derived from the particulars which are its instances and the facts that they are so. To regard them, as extreme realists do, as a primitive kind of entity, distinct from particulars but able to combine with them to yield facts, is to put the universal cart before the factual horse. It does nothing but pose such ancient but manifestly dotty conundrums as: why there are these two different kinds of

entity, particulars and universals, and what the difference between them is; why two entities of the same kind (two particulars or two universals) cannot combine to form a fact; what the relation is between a particular and a universal that are so combined. The last question on its own is fatal to this view, since any answer to it immediately generates Bradley's (1897: Ch. 3) notoriously vicious regress; a regress not avoided just by Armstrong's ingenuous device of calling the relation in question a "union . . . closer than relation" (1978a: 3).

The fact is, as Ramsey showed, that we have no *a priori* reason to suppose that universals are fundamentally different in kind from particulars. What we think of as particulars are merely the kinds of entity we can most readily individuate, typically by appeal to their spatio-temporal location (cf. Braithwaite 1926); and a universal is just the common residue of a set of facts about such individuals. So there is really no mystery about what relates particular to universal in a fact, nor about why a fact has to contain at least one of each. Nor is there any general reason why residual universals cannot themselves be individuated and so admitted in their own right as entities to be quantified over—thus, for example, leaving the particular *a* as the common residue of a set of facts about *a*'s properties. Nominalism therefore is not the only, nor the most sensible, alternative to an extreme realism about universals.

From this Ramseyan account of universals it does however follow that they need instances: Armstrong's principle of instantiation is quite right. In the law that all *F*s are *G*s, the property *F* is just the residue of such facts as *Fa*. If the law is vacuous, there are no such facts; and no facts leave no residue. If there are no *F*s, there is no *F*. So there will be no fact *FNG* to make such a vacuous law true, and the Armstrong-Dretske-Tooley theory fails. Whether there are "real connections of universals", as Ramsey put it, I do not know: like him, "I cannot deny it; for I can understand nothing by such a phrase; what we call causal laws I find to be nothing of the sort" (Ramsey 1929: 148).

7 | Epilogue

I have considered two attempts, seriously undertaken of late years, to make natural laws say more than generalizations; both fail. The law that all *F*s are *G*s is given the force it needs neither by taking it to say that 'All *F*s are *G*s' is true in all possible worlds or is in some other sense necessary, nor by taking it to assert a contingent relation between *F* and *G*. Neither construal can cover the crucial case of vacuous laws which Braithwaite rightly stresses. There are no doubt likewise aspects of the problems of laws which solutions of Braithwaite's Humean cut also have difficulty cov-

ering; only they, to my mind, are more readily patched up. Those patches, however, must be woven elsewhere.¹

■ | Notes

1. This article was written during my tenure of a Radcliffe Fellowship, for which I am indebted to the Radcliffe Trust.

■ | References

- Armstrong, D. M. 1978. *Nominalism and Realism*. Cambridge.
- Armstrong, D. M. 1978a. *A Theory of Universals*. Cambridge.
- Ayer, A. J. 1956. What is a law of nature? In his 1963 *The Concept of a Person and other essays*, pp. 209–34. London.
- Bradley, F. H. 1897. *Appearance and Reality*, 2nd edn. Oxford.
- Braithwaite, R. B. 1926. Universals and the 'method of analysis'. *Aristotelian Society Supplementary Volume* 6, 27–38.
- Braithwaite, R. B. 1953. *Scientific Explanation*. Cambridge.
- Cargile, James. 1969. The Sorites paradox. *The British Journal for the Philosophy of Science* 19, 193–202.
- Davidson, Donald. 1970. Mental events. In *Experience and Theory*, ed. L. Foster and J. W. Swanson, pp. 79–101. London.
- Dretske, Fred I. 1977. Laws of nature. *Philosophy of Science* 44, 248–68.
- Dummett, Michael. 1973. *Frege: Philosophy of Language*. London.
- Goodman, Nelson. 1965. *Fact, Fiction, and Forecast*, 2nd edn. New York.
- Hesse, Mary. 1974. *The Structure of Scientific Inference*. London.
- Hesse, Mary. 1980. A revised regularity view of scientific laws. In *Science, Belief and Behaviour*, ed. D. H. Mellor, pp. 87–103. Cambridge.
- Kneale, William. 1949. *Probability and Induction*. Oxford.
- Kripke, Saul A. 1971. Identity and necessity. In *Identity and Individuation*, ed. M. K. Munitz, pp. 135–64. New York.
- Kripke, Saul A. 1972. Naming and necessity. In *Semantics of Natural Language*, ed. Donald Davidson and Gilbert Harman, pp. 253–355; 763–9. Dordrecht.
- Lewis, David. 1973. *Counterfactuals*. Oxford.
- McGinn, Colin. 1978. Mental states, natural kinds and psychophysical laws I. *Aristotelian Society Supplementary Volume* 52, 195–220.
- Mackie, J. L. 1973. *Truth, Probability and Paradox*. Oxford.

- Mellor, D. H. 1974. In defence of dispositions. *Philosophical Review* 83, 157–81.
- Mellor, D. H. 1977. Natural kinds. *The British Journal for the Philosophy of Science* 28, 299–312.
- Mellor, D. H. 1979. The possibility of prediction. *Proceedings of the British Academy* 65.
- Mellor, D. H. 1981. McTaggart, fixity and coming true. *Reduction, Time and Reality*, ed. R. Healey. Cambridge.
- Putnam, Hilary. 1975. The meaning of 'meaning'. In his *Mind, Language and Reality*, pp. 215–71. Cambridge.
- Ramsey, F. P. 1925. Universals. In his 1978 *Foundations: Essays in Philosophy, Logic, Mathematics, and Economics*, ed. D. H. Mellor, pp. 17–39. London.
- Ramsey, F. P. 1929. General propositions and causality. In his 1978 *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*, ed. D. H. Mellor, pp. 133–51. London.
- Tooley, Michael. 1977. The nature of laws. *Canadian Journal of Philosophy* 7, 667–98.

Do the Laws of Physics State the Facts?

■ | Introduction

There is a view about laws of nature that is so deeply entrenched that it doesn't even have a name of its own. It is the view that laws of nature describe facts about reality. If we think that the facts described by a law obtain, or at least that the facts which obtain are sufficiently like those described in the law, we count the law true, or true-for-the-nonce, until further facts are discovered. I propose to call this doctrine the *facticity* view of laws. (The name is due to John Perry.)

It is customary to take the fundamental explanatory laws of physics as the ideal. Maxwell's equations, or Schrödinger's, or the equations of general relativity are paradigms, paradigms upon which all other laws—laws of chemistry, biology, thermodynamics, or particle physics—are to be modeled. But this assumption confutes the facticity view of laws. For the fundamental laws of physics do not describe true facts about reality. Rendered as descriptions of facts, they are false; amended to be true, they lose their fundamental, explanatory force.

To understand this claim, it will help to contrast biology with physics. J. J. C. Smart ([10]: chapter 3) has argued that biology . . . has no genuine laws of its own. It resembles engineering. Any general claim about a complex system, such as a radio or a living organism, will be likely to have exceptions. The generalizations of biology, or engineering's rules of thumb, are not true laws because they are not exceptionless.* If this is a

FROM *Pacific Philosophical Quarterly* 61 (1980): 75–84.

* Smart's argument for why there are no laws in biology "in the strict sense" presupposes that laws—genuine laws—have to be universal in scope. Thus, to use Smart's own example, the generalization that albinotic mice always breed true, while probably true, does not qualify as a law because the term *mouse* contains an implicit reference to a particular planet, namely, the earth. (Mice are those creatures that stand in appropriate kinship relations to animals here on earth. No

good reason [for regarding biology as a second-rate science], then it must be physics which is the second-rate science. Not only do the laws of physics have exceptions; unlike biological laws, they are not even true for the most part, or approximately true.

The view of laws with which I begin—"Laws of nature describe facts about reality"—is a pedestrian view that, I imagine, any scientific realist will hold. It supposes that laws of nature tell how objects of various kinds behave: how they behave some of the time, or all of the time, or even (if we want to prefix a necessity operator) how they must behave. What is critical is that they talk about objects—real concrete things that exist here in our material world, things like quarks, or mice, or genes; and they tell us what these objects do.

Biological laws provide good examples. For instance, here is a generalization taken from a Stanford text on chordates: (Alexander [1]:179)

The gymnotoids [American knife fish] are slender fish with enormously long anal fins, which suggest the blade of a knife of which the head is a handle. They often swim slowly with the body straight by undulating this fin. They [presumably "always" or "for the most part"] are found in Central and South America. . . . Unlike the characins they ["usually"?] hide by day under river banks or among roots, or even bury themselves in sand, emerging only at night.

The fundamental laws of physics, by contrast, do not tell what the objects in their domain do. If we try to think of them in this way, they are simply false, not only false but deemed false by the very theory which maintains them. But if physics' basic, explanatory laws do not describe how things behave, what do they do? Once we have given up facticity, I don't know what to say. Richard Feynman, in *The Character of Physical Law*, offers an idea, a metaphor. Feynman tells us "There is . . . a rhythm and a pattern between the phenomena of nature which is not apparent to the eye, but only to the eye of analysis; and it is these rhythms and patterns which we call Physical Laws . . ." ([3]: 13). Most philosophers will want to know a lot more about how these rhythms and patterns function. But at least Feynman does not claim that the laws he studies describe the facts.

creature on another planet, however mouselike, could possibly be a mouse, given the way that evolutionary biologists define species.) We could remove the implicit reference to the earth by defining *mouse* in terms of a set of properties that all mice have and which, on this planet, are possessed only by mice. But then it is quite likely that, somewhere in the vast reaches of the universe, there are animals with those properties that do not breed true. Smart concludes that "if the propositions of biology are universal in scope, then such laws are very likely not universally true. If they are not falsified by some queer species or phenomenon on earth they are very likely falsified elsewhere in the universe" (*Philosophy and Scientific Realism*, 54). Smart draws the same conclusion about psychology.

I say that the laws of physics do not provide true descriptions of reality. This sounds like an anti-realist doctrine. Indeed it is, but to describe the claim in this way may be misleading. For anti-realist views in the philosophy of science are traditionally of two kinds. Bas van Fraassen [4] is a modern advocate of one of these versions of anti-realism; Hilary Putnam ([8] [9]) of the other. Van Fraassen is a sophisticated instrumentalist. He worries about the existence of unobservable entities, or rather, about the soundness of our grounds for believing in them; and he worries about the evidence which is supposed to support our theoretical claims about how these entities behave.* But I have no quarrel with theoretical entities; and for the moment I am not concerned with how we know what they do. What is troubling me here is that our explanatory laws don't tell us what they do. It is in fact part of their explanatory role not to tell.

Hilary Putnam in his new version of transcendental realism also maintains that the laws of physics don't represent facts about reality. But this is because nothing—not even the most commonplace claim about the cookies which are burning in the oven—represents facts about reality. If anything did, Putnam would probably think that the basic equations of modern physics did best. This is the claim that I reject. I think we can allow that all sorts of statements represent facts of nature, including the generalizations one learns in biology or engineering. It is just the fundamental explanatory laws that don't truly represent. Putnam is worried about meaning and reference and how we are trapped in the circle of words. I am worried about truth and explanation, and how one excludes the other.

I | Explanation by Composition of Causes, and the Trade-Off of Truth and Explanatory Power

Let me begin with a law of physics everyone knows—the law of universal gravitation. This is the law which Feynman uses for illustration; he endorses the view that this law is “the greatest generalization achieved by the human mind” (Feynman [3]: 14).

Law of Gravitation: $F = Gmm'/r^2$.

In words, Feynman tells us ([3]: 14)

The Law of Gravitation is that two bodies exert a force between each other which varies inversely as the square of the distance between them, and varies directly as the product of their masses.

* See chapter 9 for a detailed examination of van Fraassen's brand of antirealism.

Does this law truly describe how bodies behave?

Assuredly not. Feynman himself gives one reason why. "Electricity also exerts forces inversely as the square of the distance, this time between charges . . ." ([3]: 30). It's not true that for *any* two bodies the force between them is given by the law of gravitation. Some bodies are charged bodies, and the force between them is not Gmm^1/r^2 . Rather it is some resultant of this force with the electric force which Feynman refers to.

For bodies which are both massive and charged, the law of universal gravitation and Coulomb's law (the law which gives the force between two charges) interact to determine the final force. But neither law by itself truly describes how the bodies behave. No charged objects will behave just as the law of universal gravitation says; and any massive objects will constitute a counterexample to Coulomb's law. These two laws are not true; worse, they are not even approximately true. In the interaction between the electrons and the protons of an atom, for example, the Coulomb effect swamps the gravitational one, and the force which actually occurs is very different from that described by the law of gravity.

There is an obvious rejoinder: I have not given a complete statement of these two laws, only a shorthand version. The Feynman version has an implicit *ceteris paribus* modifier in front, which I have suppressed. Speaking more carefully, the law of universal gravitation is something like this:

If there are no forces other than gravitational forces at work, then two bodies exert a force between each other which varies inversely as the square of the distance between them, and varies directly as the product of their masses.

I will allow that this law is a true law, or at least one which is held true within a given theory. But it is not a very useful law. One of the chief jobs of the law of gravity is to help explain the forces which objects experience in various complex circumstances. *This* law can explain in only very simple, or ideal circumstances. It can account for why the force is as it is when just gravity is at work; but it is of no help for cases in which both gravity and electricity matter. Once the *ceteris paribus* modifier has been attached, the law of gravity is irrelevant to the more complex and interesting situations.

This unhappy feature is characteristic of explanatory laws. I said that the fundamental laws of physics do not represent the facts, whereas biological laws and principles of engineering do. This statement is both too strong and too weak. Some laws of physics do represent facts, and some laws of biology—particularly the explanatory laws—do not. The failure of facticity does not have so much to do with the nature of physics, but rather with the nature of explanation. We think that nature is governed by a small number of simple, fundamental laws. The world is full of complex

and varied phenomena, but these are not fundamental. They arise from the interplay of simpler processes obeying the basic laws of nature.

This picture of how nature operates to produce the subtle and complicated effects we see around us is reflected in the explanations that we give: we explain complex phenomena by reducing them to their simpler components. This is not the only kind of explanation we give, but it is an important and central kind. I shall use the language of John Stuart Mill, and call this *explanation by composition of causes* (Mill [7]: Book III, Ch. VI).

It is characteristic of explanations by composition of causes that the laws they employ fail to satisfy the requirement of facticity. The force of these explanations comes from the presumption that the explanatory laws “act” in combination just as they would “act” separately. It is critical, then, that the laws cited have the same form, in or out of combination. But this is impossible if the laws are to describe the actual behavior of objects. The actual behavior is the resultant of simple laws in combination. The effect which occurs is not an effect dictated by any one of the laws separately. In order to be true in the composite case, the law must describe one effect (the effect which actually happens); but to be explanatory, it must describe another. There is a trade-off here between truth and explanatory power.

II | How Vector Addition Introduces Causal Powers

Our example, where gravity and electricity mix, is an example of the composition of forces. We know that forces add vectorially. Doesn't vector addition provide a simple and obvious answer to my worries? When gravity and electricity are both at work, two forces are produced, one in accord with Coulomb's law, the other according to the law of universal gravitation. Each law is accurate. Both the gravitational and the electric force are produced as described; the two forces then add together, vectorially, to yield the total “resultant” force.

The vector addition story is, I admit, a nice one. But it is just a metaphor. We add forces (or the numbers that represent forces) when we do calculations. Nature does not “add” forces. For the “component” forces are not there, in any but a metaphorical sense, to be added; and the laws which say they are there must also be given a metaphorical reading. Let me explain in more detail.

The vector addition story supposes that Feynman has left something out in his version of the law of gravitation. The way he writes it, it sounds as if the law describes the *resultant* force exerted between two bodies, rather than a component force—the force which is *produced between the two bodies in virtue of their gravitational masses* (or, for short, the force *due to gravity*). A better way to state the law would be

Two bodies produce a force between each other (the force due to gravity) which varies inversely as the square of the distance between them, and varies directly as the product of their masses.

Similarly, for Coulomb's law

Two charged bodies produce a force between each other (the force due to electricity) which also varies inversely as the square of the distance between them, and varies directly as the product of their charges.

These laws, I claim, do not satisfy the facticity requirement. They appear, on the face of it, to describe what bodies do: in the one case, the two bodies produce a force of size Gmm'/r^2 , in the other, they produce a force of size qq'/r^2 . But this cannot literally be so. For the force of size Gmm'/r^2 and the force of size qq'/r^2 are not real, occurrent forces. In interaction, a single force occurs—the force we call the “resultant”—and this force is neither the force due to gravity nor the electric force. On the vector addition story, the gravitational and the electric force are both produced, yet neither exists.

Mill would deny this. He thinks that in cases of the composition of causes, each separate effect does exist—it exists as *part* of the resultant effect, just as the left half of a table exists as part of the whole table. Mill's paradigm for composition of causes is mechanics. He says:

In this important class of cases of causation, one cause never, properly speaking, defeats or frustrates another; both have their full effect. If a body is propelled in two directions by two forces, one tending to drive it to the north, and the other to the east, it is caused to move in a given time exactly as far in *both* directions as the two forces would separately have carried it; . . . (Mill [7]: Book III, Ch. VI).

Mill's claim is unlikely. Events may have temporal parts, but not parts of the kind Mill describes. When a body has moved along a path due northeast, it has travelled neither due north nor due east. The first half of the motion can be a part of the total motion; but no pure north motion can be a part of a motion which always heads northeast. (We learn this from Judith Jarvis Thomson's *Acts and Other Events*.) The lesson is even clearer if the example is changed a little: a body is pulled equally in opposite directions. It doesn't budge an inch, but on Mill's picture it has been caused to move both several feet to the left and several feet to the right. I realize, however, that intuitions are strongly divided on these cases, so in the next section I will present an example for which there is no possibility for seeing the separate effects of the composed causes as part of the effect which actually occurs.

It is implausible to take the force due to gravity and the force due to

electricity literally as parts of the actually occurring force. Is there no way, then, to make sense of the story about vector addition? I think there is, but it involves giving up the facticity view of laws. We can preserve the truth of Coulomb's law and the law of gravitation by making them about something other than the facts—the laws can describe the causal powers that bodies have.

Hume taught that “the distinction, which we often make betwixt power and the *exercise* of it, is . . . without foundation” (Hume [5]: Part III, Section XIV). It is just Hume's illicit distinction that we need here: the law of gravitation claims that two bodies have the *power* to produce a force of size Gmm'/r^2 . But they don't always succeed in the *exercise* of it. What they actually produce depends on what other powers are at work, and on what compromise is finally achieved among them. This may be the way we do sometimes imagine the composition of causes. But if so, the laws we use talk not about what bodies do, but about what powers they possess.

The introduction of causal powers will not be seen as a very productive starting point in our current era of moderate empiricism. Without doubt, we do sometimes think in terms of causal powers, so it would be foolish to maintain that the facticity view must be correct and the use of causal powers a total mistake. Still, facticity cannot be given up easily. We need an account of what laws are that connects them, on the one hand, with standard scientific methods for confirming laws, and on the other, with the use they are put to for prediction, construction, and explanation. If laws of nature are presumed to describe the facts, then there are familiar, detailed philosophic stories to be told about why a sample of facts is relevant to their confirmation, and how they help provide knowledge and understanding of what happens in nature. Any alternative account of what laws of nature do and what they say must serve at least as well; and no story I know of causal powers makes a very good start.

III | A Real Example of the Composition of Causes

The ground state of the carbon atom has five distinct energy levels [see figure 7.1]. Physics texts commonly treat this phenomenon sequentially, in three stages. I shall follow the discussion of Albert Messiah in volume II of *Quantum Mechanics* [6]. In the first stage, the ground state energy is calculated by a central field approximation; and the single line (a) is derived. For some purposes, it is accurate to assume that only this level occurs. But some problems require a more accurate description. This can be provided by noticing that the central field approximation takes account only of the *average* value of the electrostatic repulsion of the inner shell electrons on the two outer electrons. This defect is remedied at the second

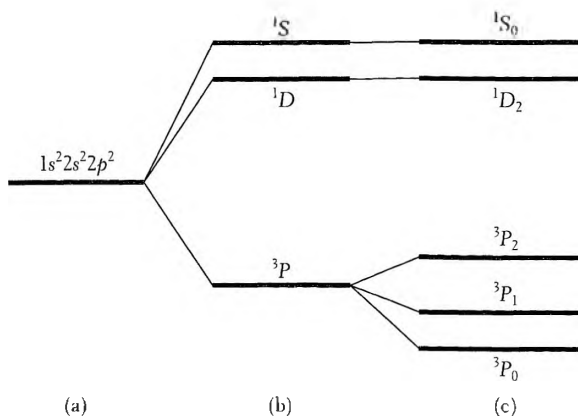


Fig. 7.1. The levels of the ground state of the carbon atom; (a) in the central field approximation ($V_1 = V_2 = 0$); (b) neglecting spin-orbit coupling ($V_2 = 0$); (c) including spin-orbit coupling. (Messiah [6])

stage by considering the effects of a term which is equal to the difference between the exact Coulomb interaction and the average potential used in stage one. This corrective potential “splits” the single line (a) into three lines depicted in (b).

Still, the treatment is inaccurate because it neglects spin effects. Each electron has a spin, or internal angular momentum, and the spin of the electron couples with its orbital angular momentum to create an additional potential. The additional potential arises because the spinning electron has an intrinsic magnetic moment, and “an electron moving in [an electrostatic] potential ‘sees’ a magnetic field” ([6]: 552). About the results of this potential Messiah tells us, “Only the 3P state is affected by the spin-orbit energy term; it gets split into three levels: 3P_0 , 3P_1 and 3P_2 .” ([6]: 706). Hence the five levels pictured in (c).

The philosophic perplexities stand out most at the last stage. The five levels are due to a combination of a Coulomb potential [which produces three energy levels], and a potential created by spin-orbit coupling [that] “splits” the lowest of these again into three. *That* is the explanation of the five levels. But how can we state the laws that it uses?

For the Coulomb effect we might try

Whenever a Coulomb potential is like that in the carbon atom, the three energy levels pictured in (b) occur.

(The real law will of course replace “like that in the carbon atom” by a mathematical description of the Coulomb potential in carbon; and similarly for “the three energy levels pictured in (b)”.) The carbon atom itself provides a counterexample to this law. It has a Coulomb potential of the right kind; yet the five levels of (c) occur, not the three levels of (b).

We might, in analogy with the vector addition treatment of composite forces, try instead

The energy levels produced by a Coulomb potential like that in the carbon atom are the three levels pictured in (b).

But (as with the forces “produced by gravity” in our earlier example) the levels which are supposed to be produced by the Coulomb potential are levels that don’t occur. In actuality, five levels occur, and they do not include the three levels of (b). In particular, as we can see from Messiah’s diagram, the lowest of the three levels—the 3P —is not identical with any of the five. In the case of the composition of motions, Mill tried to see the “component” effects as parts of the actual effect. But that certainly will not work here. The 3P level in (b) may be “split”, and hence “give rise to,” the 3P_0 , 3P_1 , and 3P_2 levels in (c); but it is certainly not a part of any of these levels.

It is hard to state a true *factual* claim about the effects of the Coulomb potential in the carbon atom. But quantum theory does guarantee that a certain *counterfactual* is true: the Coulomb potential, if it *were* the only potential at work, would produce the three levels in (b). Clearly this counterfactual bears on our explanation. But we have no model of explanation that shows how. The covering law model shows how statements of fact are relevant to explaining a phenomenon.* But how is a truth about energy levels which would occur in quite different circumstances relevant to the levels which do occur in these? We think the counterfactual is important; but we have no account of how it works.

IV | Composition of Causes Versus Explanation by Covering Law

The composition of causes is not the only method of explanation which can be employed. There are other methods, and some of these are compatible with the facticity view of laws. Standard covering law explanations are a prime example.

Sometimes these other kinds of explanation are available even when

* See Carl G. Hempel, “Two Basic Types of Scientific Explanation,” in chapter 6.

we give an explanation which tells what the component causes of a phenomenon are. For example, in the case of Coulomb's law and the law of gravity, we know how to write down a more complex law (a law with a more complex antecedent) which says exactly what happens when a system has both mass and charge. Mill thinks that such "super" laws are always available for mechanical phenomena. In fact, he thinks, "This explains why mechanics is a deductive or demonstrative science, and chemistry is not" (Mill [7]: Book III, Ch. VI).

I want to make three remarks about these super laws and the covering explanations they provide: First, super laws aren't always available; second, even when they are available, they often don't explain much; third—and most importantly—even when other good explanations are to hand, if we fail to describe the component processes that go together to produce a phenomenon, we lose a central and important part of our understanding of what makes things happen.

1 There are a good number of complex scientific phenomena which we are quite proud to be able to explain. For many of these explanations, however, super covering laws are not available to us. I argue this in "The Truth Doesn't Explain Much" (Cartwright [2]). Sometimes in these situations we have every reason to believe that a super law exists. (God has written it somewhere in the Book of Nature.) In other cases we have no good empirical reason to suppose even this much. (Nature may well be underdetermined; God failed to write laws for every complex situation.) Nevertheless, after we have seen what occurs in a specific case, we are often able to understand how various causes contributed to bring it about. We do explain, even without knowing the super laws. We need a philosophical account of laws and explanations which covers this very common scientific practice, and which shows why these explanations are good ones.

2 Sometimes super laws, even when they are available to cover a case, may not be very explanatory. This is an old complaint against the covering law model of explanation: "Why does the quail in the garden bob its head up and down in that funny way whenever it walks?" . . . "Because they all do." In the example of spin-orbit coupling it does not explain the five energy levels that appear in a particular experiment to say "All carbon atoms have five energy levels."

3 Often, of course, a covering law for the complex case will be explanatory. This is especially true when the antecedent of the law does not just piece together the particular circumstances that obtain on the occasion in question, but instead gives a more abstract description which fits with a general body of theory. In the case of spin-orbit coupling, . . . quantum mechanics provides general theorems about symmetry groups, and Hamiltonians, and degeneracies, from which we could expect to derive, covering law style,

the energy levels of carbon from the appropriate abstract characterization of its Hamiltonian, and the symmetries it exhibits.

Indeed, we can do this; and if we don't do it, we will fail to see that the pattern of levels in carbon is a particular case of a general phenomenon which reflects a deep fact about the effects of symmetries in nature. On the other hand, to do only this misses the detailed causal story of *how* the splitting of spectral lines by the removal of symmetry manages to get worked out in each particular case.

This two-faced character is a widespread feature of explanation. Even if . . . there is a single set of super laws which unifies all the complex phenomena one studies in physics . . . , our current picture may yet provide the ground for these laws: what the unified laws dictate to happen, happens *because of* the combined action of laws from separate domains, like the law of gravity and Coulomb's law. Without these laws, we would miss an essential portion of the explanatory story. Explanation by subsumption under super, unified covering laws would be no replacement for the composition of causes. It would be a complement. To understand how the consequences of the unified laws are brought about would require separate operation of the law of gravity, Coulomb's law, and so forth; and the failure of facticity for these contributory laws would still have to be faced.

V | Conclusion

There is a simple, straightforward view of laws of nature which is suggested by scientific realism, the facticity view: laws of nature describe *how* physical systems behave. This is by far the commonest view, and a sensible one; but it doesn't work. It doesn't fit explanatory laws, like the fundamental laws of physics. Some other view is needed if we are to account for the use of laws in explanation; and I don't see any obvious candidate which is consistent with the realist's reasonable demand that laws describe reality and state facts which might well be true. There is, I have argued, a trade-off between factual content and explanatory power. We explain certain complex phenomena to be the result of the interplay of simple, fundamental laws. But what do these fundamental laws *say*? To play the role in explanation we demand of them, these laws must have the same form when they act together as when they act singly. In the simplest case, the consequences which the laws prescribe must be exactly the same in interaction, as the consequences which would obtain if the law were operating alone. But then, what the law states cannot literally be true, for the consequences which would occur if it acted alone are not the consequences which actually occur when it acts in combination.

If we state the fundamental laws as laws about what happens when only a single cause is at work, then we can suppose the law to provide a true description. The problem arises when we try to take that law and use it to explain the very different things which happen when several causes are at work. This is the point of "The Truth Doesn't Explain Much" (Cartwright [2]). There is no difficulty writing down laws which we suppose to be true: "If there are no charges, no nuclear forces, . . . then the force between two masses of size m and m' separated by a distance r is Gmm'/r^2 ." We count this law true—what it says will happen, does happen—or at least happens to within a good approximation. But this law doesn't explain much. It is irrelevant to cases where there are electric or nuclear forces at work. The laws of physics, I concluded, to the extent that they are true, don't explain much. We could know all the true laws of nature, and still not know how to explain composite cases. Explanation must rely on something other than law.

But this view is absurd. There aren't two vehicles for explanation—laws for the rare occasions when causes occur separately; and another secret, nameless device for when they occur in combination. Explanations work in the same way whether one cause is at work, or many. "Truth Doesn't Explain" raises perplexities about explanation by composition of causes; and it concludes that explanation is a very peculiar scientific activity, which commonly does not make use of laws of nature. But scientific explanations do use laws. It is the laws themselves which are peculiar. The lesson to be learned is that the laws which explain by composition of causes fail to satisfy the facticity requirement. If the laws of physics are to explain how phenomena are brought about, they cannot state the facts.¹

■ | Notes

1. This paper was given as part of a symposium on "Explanation and Scientific Realism" at the School of Philosophy, University of Southern California, March 1980.

■ | References

- [1] Alexander, R. McNeill. *The Chordates* (Cambridge University Press, 1975).
- [2] Cartwright, Nancy. "The Truth Doesn't Explain Much". *American Philosophical Quarterly*, 17 (1980). [Reprinted in N. Cartwright, *How the Laws of Physics Lie* (Oxford: Clarendon Press, 1983).]
- [3] Feynman, Richard. *The Character of Physical Law*. (Cambridge, Mass.: MIT Press, 1967).
- [4] van Fraassen, Bas. *The Scientific Image* (Oxford University Press, 1980).

- [5] Hume, David. *A Treatise of Human Nature* (edited by L.A. Selby Bigge. Oxford: Clarendon Press, 1978).
- [6] Messiah, Albert. *Quantum Mechanics* (Amsterdam: North Holland, 1961).
- [7] Mill, John Stuart. *A System of Logic* (London: John W. Parker and Son, 1856).
- [8] Putnam, Hilary. *Meaning and the Moral Sciences*. (London: Routledge and Kegan Paul, 1978).
- [9] Putnam, Hilary. "Models and Reality." [In H. Putnam, *Realism and Reason*, vol. 3, *Philosophical Papers* (Cambridge: Cambridge University Press, 1983), pp. 1-25.].
- [10] Smart, J. J. C. *Philosophy and Scientific Realism* (London: Routledge and Kegan Paul, 1963).

- 7.1 The Regularity Theory: Ayer and Hume** 879
 Hume on Causation 879
 The Problem of Vacuous Laws 880
 The Problem of Noninstantial Laws 881
 The Missing-Values Problem for Functional Laws 882
 The Problem of Accidental Generalizations 883
- 7.2 Ayer's Epistemic Regularity Theory** 883
 Dretske's Criticism of Ayer's Theory 884
- 7.3 Dretske's Universals Theory** 885
 Extensions and Intensions 886
 Dretske's Necessitarian View 886
 The Identification Problem 887
 The Inference Problem 888
 Mellor's Criticism of Dretske: Can Laws Be Noninstantial? 890
- 7.4 Mellor's Defense of the Regularity Theory** 891
 Kripke and Putnam on Natural Kinds and Essences 892
 Putnam's New Theory of Reference 893
 Mellor's Criticisms of Putnam's Theory of Reference 894
- 7.5 Cartwright's Antirealism about Fundamental Laws** 896
 A Response to Cartwright's Antirealist Argument 897
- 7.6 Summary** 898

7.1 | The Regularity Theory: Ayer and Hume

A. J. Ayer defends a version of the regularity theory of laws in his article, "What Is a Law of Nature?" As Ayer notes, the regularity theory has its origins in David Hume's analysis of causation.

HUME ON CAUSATION

In his *Treatise of Human Nature* (1739), David Hume (1711–76) advocated what is called the *constant conjunction* (or *regularity*) *theory of causation*. According to Hume, the claim that one event, *a* (of type *A*), caused another event, *b* (of type *B*), means only that *A*-events are, as a matter of empirical fact, always followed by *B*-events. (From now on, we shall refer to events of type *A* simply as *As* and to events of type *B* as *Bs*.) Objectively speaking, the causal relation between *a* and *b* is nothing more than the constant conjunction of *As* and *Bs*. If *As* always have been and always will be followed by *Bs*, then *As* cause *Bs* and, in particular, *a* caused *b*. Hume denied any objective necessity "out there in the world" between *As* and *Bs* in virtue of which *A*-events produce *B*-events or make *B*-events occur. Our conviction that effects do not merely happen to follow causes but in some sense must necessarily occur given the appropriate cause results from a purely subjective feeling in our minds when we experience an *A* (or imagine an *A*) and expect a *B* to follow. Thus, on Hume's view, we regard *As* as the cause of *Bs* when past experience has induced in us the expectation that *As* will always be followed by *Bs* in the future.

Hume's constant-conjunction theory stands our commonsense view of causation on its head. For, typically, we think that we first discover the fact that *As* causally necessitate *Bs*, and then, on the basis of that discovery, understand why *As* have always been followed by *Bs* and predict that *As* always will be followed by *Bs*. As Ayer explains, part of the case for Hume's radical reversal of our usual thinking about causality rests on his skeptical, empiricist analysis of the idea of necessity. If there is a necessary connection between causes and their effects, then the necessity is either logical or nonlogical. Hume denies that the connection can be logical. If it were logical, then effects could be deduced from causes and we could know, prior to experience, that one kind of event (the cause) must invariably be associated with another kind of event (the effect). But no such deduction and no such a priori knowledge of effects is possible. For any cause, it is logically possible that its usual effect not follow it; moreover, our knowledge of causal relations is derived solely from experience. Thus, Hume

and Ayer argue that the kind of necessity involved in causation cannot be logical.

Curiously, Ayer does not then consider Hume's case against the second alternative, namely that some kind of nonlogical necessity links causes with their effects. Hume's argument here is epistemological.¹ He considers the two sources most commonly claimed as the origin of our idea of causal necessity—namely, perception and the will—and argues that neither gives us any experience of necessity. When Hume examines his perceptual experiences, he cannot find in them any element (what Hume calls “an impression”) of necessity. For example, when one billiard ball collides with another, what we literally see, according to Hume, is simply one motion followed by another, not the first ball making the second ball move. With regard to the willing of our actions, Hume denies that we know, independently of experience, which acts of will must be followed by which actions. Indeed, we are completely ignorant of the immediate effects of willing (presumably it is some change in the brain that then causes impulses to be transmitted by our nerves to our muscles), and only by experience do we learn which parts of our bodies we can control and which we cannot. As the possibility of sudden paralysis demonstrates, there is no logical or physical guarantee that a particular act of willing will be followed by a particular motion of one's body.

Having failed to locate any impression of necessity in either our perceptual experiences or in our ability to will actions, Hume concludes that the source of our idea of nonlogical necessity must be purely subjective. Hence, Hume sees nonlogical necessity as an imaginative fiction, something originating from our patterns of inference and expectation, which we then project onto nature and mistake for something objective.

So much by way of background in Hume's theory of causation: let us return to laws. The regularity theory of laws is often called the *Humean theory* because it, too, denies that any sort of objective nonlogical necessity connects the items appearing in a law. In its simplest form, the regularity theory says that laws of nature are nothing more than true universal generalizations. If it is a law that all copper conducts electricity, then what makes it a law is the fact that all pieces of copper, past, present, and future, conduct electricity. If it is a law that all metals expand when heated, then it is a law because, as a matter of fact, that is how all metals always behave. According to the regularity theory, the objective content of laws is exhausted by what actually happens in the world.

THE PROBLEM OF VACUOUS LAWS

There are a number of serious problems with the simple version of the regularity theory of laws. These problems are discussed by Ayer, and briefly by Dretske and Mellor, in the readings in this chapter. First, there is the problem of avoiding what Ayer calls “vacuous laws.” Modern logicians

regard the generalization "All As are Bs" as logically equivalent to "It is false that there is an A that is not a B." Thus, "All copper conducts electricity" is true if and only if it is not the case that there is a piece of copper that is not a conductor. If we translate the same generalization using predicate logic, then we get $(x)(Cx \supset Ex)$. Literally, for any x , if x is copper, then x conducts electricity. But just as before, this generalization is logically equivalent to the statement "it is not the case that there is a piece of copper that is not a conductor." In symbols we would write $\sim(\exists x)(Cx \& \sim Ex)$: it is not the case that there exists any x , such that x is copper and x is not a conductor. Now the important point to notice is that if, as a matter of fact, there is no copper in the universe, then it automatically becomes true that it is not the case that there is a piece of copper that is not a conductor. But this last statement is logically equivalent to the generalization "All copper conducts electricity." According to the simple version of the regularity theory, laws are true universal generalizations. Thus, any generalization that is automatically true simply because there are no instances of its antecedent is a law. For example, according to the simple version of the regularity theory, it is a law that all perpetual motion machines weigh ten tons, that all mermaids contain chlorophyll, and that all particles traveling faster than light are red. Clearly, this is absurd. How can the regularity theorist avoid the problem of vacuous laws?

The most obvious response is to add a further condition, an existential condition, to the regularity analysis: a true universal generalization is a law provided there actually are objects satisfying the generalization. In symbols, "All Cs are Es" is a law if and only if $(x)(Cx \supset Ex) \& (\exists x)(Cx)$. This eliminates the counterexamples mentioned in the previous paragraph, since there are no mermaids, perpetual motion machines, or particles traveling faster than light. Hence, on the amended regularity analysis, no vacuously true generalizations will qualify as laws.

THE PROBLEM OF NONINSTANTIAL LAWS

Despite the advantage of modifying the regularity theory to avoid the problem of vacuous laws, Ayer and many other regularity theorists argue that the existential condition is too strong, because it rules out some of the most important laws in science. The classic example is Newton's first law of motion, which states the principle of rectilinear inertia:

All bodies on which no net external force is acting either remain at rest or move at uniform velocity in a straight line.

Now it seems reasonable to assume that, since all bodies exert a gravitational pull on all the other bodies in the universe, no bodies are ever free from net external forces. Nonetheless, scientists accept Newton's first law not because it lacks instances but because it expresses an important truth

about the world. Newton's first law is thus an example of a nonvacuous but noninstantial law. Here are some other examples of nonvacuous non-instantial laws: if two perfectly elastic bodies were to collide, the total kinetic energy of the system would be the same before and after the impact (discussed by C. D. Broad in the article referred to by Ayer); in a perfectly reversible process, the entropy remains constant; all lumps of plutonium weighing more than one million tons conduct electricity.

The natural inclination is to handle noninstantial laws in terms of subjunctive or counterfactual conditionals, that is, in terms of how objects of a certain kind would behave if there were such objects. But this approach does not seem to be open to the regularity theorist. For regularity theorists take statements of law to describe how actual objects behave, not how possible objects would behave if, contrary to fact, they were to exist.

Broad's proposal (reported by Ayer) for reconciling noninstantial laws with the regularity theory suggests that we distinguish ultimate laws of nature from derivative laws. Thus, all ultimate laws are taken to be instantial, but laws derived from one or more such ultimate laws may not be. Consider, for example, Newton's second law:

If a net force, F , acts on a body of mass, m , then the body experiences an acceleration, $a = F/m$.

We can derive Newton's first law from his second law on the supposition that no net force is acting on a body. For according to Newton's second law, if the net force on a body were zero, then the acceleration of the body would also be zero. So Newton's first law is a noninstantial, derivative law because it can be derived from the instantial, ultimate second law. Similarly, there are ultimate, instantial laws of impact and motion which entail the nonvacuous, noninstantial law that, if two perfectly elastic bodies were to collide, kinetic energy would be conserved. But as Ayer notes, even if we could always find ultimate laws that would reconcile noninstantial laws with the regularity theory, that theory would still encounter severe problems with what are sometimes called *functional laws*.

THE MISSING-VALUES PROBLEM FOR FUNCTIONAL LAWS

Functional laws assert a functional relation between two or more variables in the form of a mathematical equation. For example, Hooke's law, $F = kx$, says that the force, F , exerted by a spring is directly proportional to x , the amount the spring is stretched. Similarly, Hubble's law $V = HD$, says that the velocity, V , with which galaxies are moving away from each other is directly proportional to D , their distance apart. The ideal gas law, $PV = nRT$, asserts that the pressure times the volume of n moles of gas is proportional to the absolute temperature of the gas. In all these functional laws, the magnitude of the variables range over an infinite number of

values only a small finite number of which will ever be realized. For example, no gas will actually be heated to all possible temperatures, nor will every spring be stretched to all possible lengths. Nonetheless, the ideal gas law tells us what the pressure of the gas would be at a temperature of 1 million degrees, and Hooke's law says what force a spring would exert if it were stretched to a hundred times its normal length. Thus, the missing values problem leads us inexorably to using subjunctive conditionals to express the content of laws in counterfactual situations. Once again, the problem for the regularity theorist is making sense of these counterfactual conditionals while still regarding laws as descriptions of what actually happens in the world.

THE PROBLEM OF ACCIDENTAL GENERALIZATIONS

Closely related to the missing-values problem for functional laws is the more general difficulty of distinguishing between genuine laws and so-called accidental generalizations (what Ayer calls "generalizations of fact" as contrasted with "generalizations of law"). Consider one of Fred Dretske's examples. Suppose that, as a matter of brute fact, the only dogs that have been or ever will be born at sea are cocker spaniels. Thus, "all dogs born at sea are cocker spaniels" emerges as a true, universal generalization. But, clearly, we would not on this basis predict that, if the dog on board our ship were a dachshund, then she would give birth to cocker spaniel puppies. Our expectations about this and other counterfactual situations depend, not on accidental generalizations, but on genuinely lawful ones. We rely on the biological law that purebred dogs produce dogs of the same breed (at least, when mated with dogs of the same breed). But the simple version of the regularity theory cannot distinguish between those universal generalizations that are laws and those that are not. So the simple version of the regularity theory is inadequate.

7.2 | Ayer's Epistemic Regularity Theory

Because of the difficulty in distinguishing between laws and accidental generalizations, most proponents of the regularity theory advocate a more sophisticated version of the theory according to which laws are true, universal generalizations with some additional features. As Ayer puts it, "the difference between our two types of generalization lies not so much on the side of the facts which make them true or false, as in the attitude of those who put them forward" (822). In his paper "Laws of Nature," Dretske summarizes the sophisticated version of the regularity theory with the formula

law = universal truth + X,

where the usual candidates for X include:

- our willingness to use the generalization in question to make predictions, especially about counterfactual situations;
- our acceptance of the generalization as well confirmed even though we have examined only a relatively small, finite number of its instances;
- the role that the generalization plays in a deductively organized system of (scientific) statements; and
- our recognition that the generalization (unlike a mere generalization of fact) explains its instances.

Because all of these candidates for X involve our beliefs and epistemic attitudes, this sophisticated version of the regularity theory is often called the *epistemic regularity theory*. Ayer's own proposal falls into this category. In Ayer's own words:

Accordingly I suggest that for someone to treat a statement of the form 'if anything has Φ it has Ψ ' as expressing a law of nature, it is sufficient (i) that subject to a willingness to explain away exceptions he believes that in a non-trivial sense everything which in fact has Φ has Ψ (ii) that his belief that something which has Φ has Ψ is not liable to be weakened by the discovery that the object in question also has some other property X, provided (a) that X does not logically entail not- Ψ (b) that X is not a manifestation of not- Ψ (c) that the discovery that something had X would not in itself seriously weaken his belief that it had Φ (d) that he does not regard the statement 'if anything has Φ and not-X it has Ψ ' as a more exact statement of the generalization that he was intending to express. (824)

DRETSKE'S CRITICISM OF AYER'S THEORY

As Ayer himself acknowledges (in his concluding paragraph), his proposal completely ignores the missing-values problem for functional laws. Even more striking is Ayer's candid admission that his proposal cannot be construed as an attempt to define what laws are. A definition of a concept would give both necessary and sufficient conditions, but Ayer offers only sufficient conditions for lawlikeness. In other words, Ayer recognizes that many things could turn out to be laws, even though they fail to satisfy his conditions. As he says, "I do not claim that to say that some proposition expresses a law of nature entails saying that someone has a certain attitude towards it; for clearly it makes sense to say that there are laws of nature which remain unknown" (824).

Ayer's candid admission about the limitations of his epistemic regularity analysis of laws—that is, its inability to countenance the existence of unknown laws—is a powerful objection to the whole approach. Dretske thinks the objection is decisive. As long as *X* includes factors that refer essentially to human beliefs, attitudes, and practices, the epistemic regularity approach entails that there are no unknown laws. As Dretske sees it, the epistemic regularity approach has confused an epistemological issue (why we believe something is a law of nature) with an ontological issue (what sort of thing a law of nature is). A more promising approach, in Dretske's view, is to address the ontological issue directly. First we should understand what laws of nature are and then (but only then) explain why we adopt towards them the attitudes that we do.

7.3 | Dretske's Universals Theory

Dretske proposes a necessitarian analysis of laws. (Similar proposals have been defended by D. M. Armstrong and Michael Tooley.) Dretske thinks that the law that we would usually express by saying "All *F*s are *G*" really has the form

$$F\text{-ness} \rightarrow G\text{-ness},$$

where *F*-ness and *G*-ness are the properties of being *F* and *G*. The term *F*-ness refers to a universal, the property a thing must have in order to be *F*. Dretske suggests that we read the connective " \rightarrow " as "yields" or "brings with it." (Tooley calls it "nomic necessitation," and Armstrong usually calls it "necessitation.") The Dretske-Armstrong-Tooley approach is called the *universals theory* because it regards laws of nature as being fundamentally about relations between universals (properties). Statements of laws of nature, on this view, are not universal generalizations about particulars but singular statements about universals.

Notice here one *prima facie* advantage the universals theory might, on a certain view of universals, be thought to have over the regularity theory. Adherents of the regularity view had difficulty explaining why laws, which are taken only to describe the way objects are, nevertheless support their counterfactuals. On the universals theory laws can support contrary-to-fact possibilities because universals are taken to be properties that can be variously possessed, or not, by objects that do, or could, exist. Given a law expressing (for example) the relation between electrical charge and magnetic field, we might reasonably go on to speak of the magnetic properties of vertebrates if they were electrically charged; likewise, if laws do indeed express relations among universals, we might reasonably state how a body would behave if, contrary to fact, no net force were acting on it.

EXTENSIONS AND INTENSIONS

One of the keys to understanding the universals theory of laws is to appreciate the difference between the *extension* and the *intension* of a predicate. A predicate is any term that, like an adjective, can be used to describe a thing. For example, the words *cat*, *elastic*, and *copper* are all predicates. The extension of a predicate is the set of objects (animals, regions of space) to which the term correctly applies. For example, the extension of *cat* is the set of all the objects that the term *cat* denotes, namely all the cats in the world.

It is more difficult to explain what the intension of a term is, and philosophers have differed in their accounts of it. The basic idea is that the intension of a predicate is its meaning (or what Mill called its *connotation*). The intension of *cat* is whatever the term *cat* means—the property of felinity if you like or, perhaps, the concept of catness. It is evident that two terms can mean different things (i.e., have different intensions) but apply to exactly the same set of objects (i.e., have the same extension). For example, all mammals (even whales and porpoises) have hair somewhere on their bodies, and mammals are the only animals that have hair. So the terms *mammal* and *hairy* (in the sense of having at least some hair) are coextensive, since they pick out the same group of objects. But clearly the intensions of these terms are different. Even though all and only mammals have hair, the term *mammal* does not mean *hairy*.

Now, suppose that it is a law of nature, a biological law, that all mammals have mammary glands. (It is biologically necessary for mammals, which suckle their young, to have milk-secreting glands.) If it is a universal truth that all mammals have mammary glands, then it must also be a universal truth that all Ys have mammary glands, where Y is any term that is coextensive with *mammal*. So, in particular, it is a universal truth that all hairy animals have mammary glands. But, as Dretske points out, from the fact that “All mammals have mammary glands” is a law and *hair* is coextensive with *mammal*, it does not follow that it is a law that all hairy animals have mammary glands. Laws, on Dretske’s account, are *opaque* in a way that normal statements about universals may not be: statements that we can deduce from laws by the substitution of terms will themselves be laws only if the terms in question have the same intension. As Dretske puts it, “laws imply universal truths, but universal truths do not imply laws” (830).

DRETSKE’S NECESSITARIAN VIEW

One of the subtleties of Dretske’s universals theory, which distinguishes it from earlier necessitarian analyses of laws, is the contrast he draws between treating laws as intensional relations between extensions and treating them as extensional relations between intensions. Many necessitarians in the

past have espoused the first position. That is, they have supposed that laws are opaque because of the special, intensional, nature of the relation connecting the extensions of terms in statements of laws. Thus, on this older necessitarian view, the reason for the lawlike character of "all copper conducts electricity" is the special strong relation of (nonlogical) necessitation connecting things that are copper to things that conduct electricity. On Dretske's alternative view, there is no need for such a special strong relation of necessitation. The items linked by the law are not physical objects or events but properties (universals, intensions). According to Dretske, a law asserts that one property (an intension) is invariably associated with another property (another intension). As long as one were to substitute another term that picks out the same property in a lawlike statement, the new statement thus generated would also be a law. When we construe terms intensionally (as denoting properties), they will be coextensive whenever they denote the same property. Thus, on Dretske's view, the relation symbolized by " \rightarrow " is extensional, while the things linked by the relation (namely properties, universals) are intensional.

We can summarize the difference between the two brands of necessitarianism by contrasting how they would represent the law that all F s are G . Earlier necessitarians would write, " $(x)(Fx \text{ [N]} \rightarrow Gx)$ "—to be read as, "anything that is F must (in some nonlogical, physical sense of *must*) be G ." Dretske would write, " $F\text{-ness} \rightarrow G\text{-ness}$ "—to be read as, "the property of being an F necessitates the property of being a G ."

There are several problems with Dretske's universals theory of laws, to which critics such as Bas van Fraassen and D. H. Mellor have drawn attention. Van Fraassen calls two of these problems the *identification problem* and the *inference problem*. The identification problem is the challenge—posed to defenders of the universals theory—of giving an adequate account of the necessitation relation that allegedly holds between those universals comprising a law. The second (and related) inference problem concerns the inferential relation between laws and their instances. The universals theorist insists that not only does the FG -law logically imply that all F s are G , but also that the FG -law explains the "mustness" or necessity that we think holds between the particular things that are instances of the law. Presumably, then, the FG -law implies either that it is necessary that all F s are G or that if some particular thing is F then it must also be G . As we shall see, it is difficult to make sense of either inference if, as the universals theory supposes, laws of nature are themselves not necessary but contingent.

THE IDENTIFICATION PROBLEM

The identification problem is that of giving an account of the necessitation relation that, according to theories such as Dretske's, connects the universals that make up a law. Recall from our earlier discussion that empir-

icists such as Hume deny that we can make sense of any notion of objective necessity differing from logical necessity. As far as Hume was concerned, nomic necessity is simply subjective, a feeling; there is no objective, nonlogical necessity that connects objects, events, or universals. What if Hume were right about this? Does this rule out the possibility that laws involve a logically necessary connection between universals?

If the necessitation relation between the universals in a law were logical, then all laws of nature would be logically necessary. In that case, if "All *F*s are *G*" is a law, it would be logically impossible for there to be an *F* that was not a *G*. To many empiricist philosophers, this is a sufficient reason to reject reading nomic necessity as logical necessity, since we usually assume that laws of nature are contingent, not necessary, truths. But a degree of caution is advisable here. As Mellor points out in his "Necessities and Universals in Natural Laws," some philosophers (notably Kripke and Putnam) have argued that, although all the laws of empirical science are discovered through empirical research, nonetheless many of those laws are not contingent but necessary. This issue is explored later in this commentary, in the section "Mellor's Defense of the Regularity Theory."

THE INFERENCE PROBLEM

The inference problem is that of explaining the "mustness," or necessity, that necessitarians believe to hold between the particular things or events that are instances of a law. For example, if necessitarians believe that it is a law that gold has an atomic number of 79 and that a particular piece of metal is gold, then they would infer that the piece of metal in question *must* have an atomic number of 79. As Dretske acknowledges, one of the main challenges for his universals theory of laws is to explain the "mustness" that appears in the conclusion of this kind of inference. On Dretske's theory, we can write out the inference as follows:

$$\begin{array}{l} F\text{-ness} \rightarrow G\text{-ness} \\ a \text{ is } F \\ \hline a \text{ must be } G. \end{array}$$

Where does the *must* in the conclusion come from? Presumably, it derives from some *must* implicit in the first premise, which is Dretske's way of representing the law that all *F*s are *G*. In ordinary English, we could state the law using *must* in two different ways. We could say either "It must be the case that, if anything is *F*, then it is *G*" or "Anything that is *F* must be *G*."

Clearly, the first way of stating the law will not help. From the premises "It must be the case that, if *x* is *F*, then *x* is *G*" and "*a* is *F*," we

cannot validly infer the conclusion that *a* must be *G*. All that follows is that, as matter of contingent fact, *a* is *G*.²

So we have to read the law as Dretske intends it to be read, as saying "Anything that is *F* must be *G*." But as far as Dretske is concerned, this is just another way of saying that the property of being *F* necessitates the property of being *G*. In other words, according to Dretske's theory, the law thus stated is not a generalization about objects or events (which are particulars); it is a singular statement about properties (which are universals). To repeat, on Dretske's view, the law does not say "Each individual thing that is *F* must also be *G*." Rather, it says that the property of *F*-ness necessitates, or brings with it, the property of *G*-ness. How, then, given Dretske's understanding of what the *FG*-law asserts, can we use it to deduce the conclusion that a particular thing must be *G*? As Dretske admits, the only inference that is uncontroversially valid is the inference from "*F*-ness \rightarrow *G*-ness" to "All *F*s are *G*." But "All *F*s are *G*" does not validly imply that anything that is *F* must also be *G*. (Were this inference valid, the distinction between accidental generalizations and laws would collapse.)

In response to this difficulty, Dretske offers an analogy with the offices and branches of a government. There is a legal code in the United States that lays down what powers pertain to the office of the president, the two houses of Congress, and the Supreme Court, and how these branches of government are related to one another. The code itself is contingent; the Constitution of the United States could have been different. But given that Constitution, it is now true of anyone who holds the office of president that that person must consult Congress and receive its approval before declaring war. The law is not about the particular people who hold the various offices; the law is about the powers and duties of the offices themselves and the relations between them. But because the law is what it is, anyone who holds a particular office *must* behave in a certain way.

An analogy is not a proof, and Dretske does not claim to have proven anything. Some critics (such as van Fraassen) have stressed the disanalogies between a legal code and laws of nature. With a legal code, we understand the origin of the law's prescriptive force, namely the commitment of citizens to enforce it. But what is the origin of the analogously prescriptive force of laws of nature? Given the mystery that seems to surround the notion of contingent relations between universals, other advocates of the universals theory have candidly admitted that necessitation is an inexplicable basic concept that the theory is forced to postulate. The justification for accepting the concept lies in the superiority claimed for the universals theory as a whole in accounting for the essential features of laws.

MELLOR'S CRITICISM OF DRETSKE: CAN LAWS BE NONINSTANTIAL?

Recall the *prima facie* advantage discussed earlier that the universals theory might be thought to have over the regularity theory: that is, if universals can exist without being instantiated, then noninstantial laws can be accommodated easily by the universals theory but not so easily by the regularity theory. Mellor argues that this supposed advantage is illusory because the universals involved in laws must, after all, have instances. He reaches this conclusion by adopting the account of universals championed by the British mathematician and philosopher Frank Plumpton Ramsey (1903–30).

To appreciate the motivation for Ramsey's proposal, consider the traditional view that there are particulars—individual objects such as *this* apple and *that* apple—which exist in space and time—and universals—properties such as redness and greenness—which are neither spatial nor temporal. Somehow or other, on the traditional view, these two very different kinds of entity—particular objects and universal properties—combine to form facts, such as the fact that this apple is red. But how can this combination be possible given that particulars and universals are so utterly dissimilar? Taking his cue from Wittgenstein's *Tractatus*, Ramsey proposed that the world consists, not of particular things and universal properties, but of facts.³ According to Ramsey, universals and particulars should be regarded, not as two fundamentally different sorts of thing each having an independent existence, but as mutually necessary parts of particular facts. On this view (which is a version of nominalism), particulars and universals alike are aspects of facts. For example, the particular denoted by *this apple* is what is common to all the particular facts about this apple, while the universal denoted by *red* is what is common to all the particular facts about red things.

Mellor agrees with Ramsey that if we make the mistake of thinking of universals and particulars as having some kind of independent real existence, then we will be led into such “dotty conundrums” (861) as worrying about how two such radically different kinds of thing can combine to form a fact. On Ramsey's view, universals are just the part that is common to all the facts of the relevant class. The relevant class of facts for the *FG-law* must include facts such as *Fa*. Therefore, *F-ness*, the universal in the *FG-law*, must have instances if the law is to be genuine. Genuine laws, on such an account of universals, cannot be non-instantial.

7.4 | Mellor's Defense of the Regularity Theory

In "Necessities and Universals in Natural Laws," Mellor defends the Humean regularity theory of laws. A key element of that theory is the insistence that laws of nature are contingent, not necessary, truths, and this insistence is one of the main differences between the regularity theory and traditional versions of the necessitarian theory. Until quite recently, it was thought that, whatever the defects of the regularity theory as an adequate account of laws, it was at least correct about the contingency of laws. And as we have seen, the conviction that laws of nature are contingent has proven to be a stumbling block for the newer, universals version of the necessitarian theory.⁴

The main arguments for the contingency of laws derive from Hume. Take any scientific law of the form "All *F*s are *G*." Even though the law is true, we can easily conceive that something could be *F* without that thing also being *G*. Since whatever is conceivable is possible, the *FG*-law could be false. Thus, it and all other similar laws are contingent. Moreover, laws of nature are discovered by empirical research, and this seems to be the only way we can find out which lawlike generalizations are true and which false. If laws of nature were necessary, then we should be able to discover them through a priori reasoning. But we cannot do this. Therefore, the laws of nature are contingent.

It is now widely acknowledged that Hume's arguments are unsound. Conceivability is not an infallible guide to possibility.⁵ Moreover, as Saul Kripke has demonstrated, the fact that a proposition is a posteriori does not entail that it is contingent.⁶ The classic example is the simple identity claim that Hesperus is (that is, is identical with) Phosphorus. The discovery that the two names *Hesperus* and *Phosphorus* refer to one and the same physical object (sometimes seen around sunset, at other times seen around sunrise) was empirical. No amount of a priori reasoning about the meanings of the names *Hesperus* and *Phosphorus* could have revealed this truth to us. Nonetheless, the fact that the object picked out by the name *Hesperus* (namely, the planet Venus) is identical with the object picked out by the name *Phosphorus* (namely, the planet Venus) is a necessary truth, since every object is necessarily identical with itself.

In understanding Kripke's view about the necessity of identity claims, it is important to appreciate the distinction between metaphysical necessity and logical necessity. As with logical necessity, it is impossible for a metaphysically necessary proposition to be false (that is, there is no possible world in which such a proposition is false). But unlike a logically necessary proposition, its necessary truth is not guaranteed solely by logic and definitions. The proposition that Hesperus is identical with Phosphorus is metaphysically necessary, but not logically necessary.

Kripke's case for the necessity of identity claims rests on his theory of

reference, specifically his theory of rigid designation. Obviously, if it were possible for the names *Hesperus* and *Phosphorus* to refer to different planets, then the assertion that Hesperus is identical with Phosphorus would not be necessary. By saying that the names *Hesperus* and *Phosphorus* are rigid designators, Kripke is denying that this is possible. There is not time here to consider Kripke's ingenious (and, to many minds, convincing) arguments for his thesis that simple names such as *Hesperus* and *Phosphorus* are rigid designators.⁷ Suffice it to say that part of the plausibility of his thesis rests on the fact that names such as *Hesperus* and *Phosphorus* are indeed simple names with no connotative meaning. So it is tempting to think that their reference, the things they refer to by virtue of some initial baptism, completely exhausts whatever meaning they have.

Of immediate concern to us is whether something like Kripke's view can be extended to laws of nature. The problem is that only a few scientific laws make identity claims, and when they do, those claims concern classes of objects and properties, rather than single objects designated by simple names. Saul Kripke and Hilary Putnam have argued that, as with the identity of Hesperus and Phosphorus, a significant class of laws of nature are metaphysically necessary, namely, those laws attributing essential properties to natural kinds. It is these arguments that Mellor tries to refute on behalf of the regularity theory.

KRIPKE AND PUTNAM ON NATURAL KINDS AND ESSENCES

Natural kinds have traditionally been thought to include such things as chemical elements and compounds like arsenic and hydrochloric acid, and biological species like tigers and elm trees. The basic idea is that each member of such classes shares a common nature in virtue of which it belongs to that relevant kind. The intended contrast is with artificial kinds or groups, such as all animals weighing over fifty pounds or all compounds whose chemical name in English begins with the letter A, where there is nothing else that the items in these groups need have in common. Essential properties are those properties of a thing that it cannot exist without—or, perhaps, properties that a thing of some kind cannot lack while remaining of that kind. If the essential properties of tigers are P_1 , P_2 , and P_3 , then nothing that lacks one or more of these properties can belong to the natural kind *tiger*. Thus, the doctrine of natural kinds and the notion of essential properties go hand in hand: natural kinds are classes of things sharing a common nature or core set of essential properties.

Kripke and Putnam usually take the essential properties of things to be their microstructural properties. They say, for example, that being H_2O is an essential property of water; having atomic number 79 is an essential property of gold; having a particular set of genes is essential for membership in a biological species. Consider the example of gold. Undoubtedly, modern scientists believe that having atomic number 79 is a fundamental

property of gold. According to Kripke and Putnam, gold is, by its very nature, something that has 79 protons in its nucleus and 79 orbiting electrons. It is possible that some pieces of gold may not be hard, shiny, or yellow, but it is impossible that any piece of gold—any piece of that very element—could fail to have atomic number 79. So according to Kripke and Putnam, the law that gold has an atomic number of 79 is a necessary truth: given the essential nature of gold, it could not be otherwise.⁸ Of course, the discovery of the law is empirical, but it does not follow from this that the law is contingent; statements can be necessary and yet a posteriori. Consider our earlier example that *Hesperus* is identical with *Phosphorus*. Given what *Hesperus* and *Phosphorus* designate (namely, the planet Venus), the identity could not fail to hold, since one thing could not be two distinct things; nonetheless, it was an empirical discovery that *Hesperus* and *Phosphorus* name one and the same planet.

Mellor rejects the Kripke-Putnam doctrine of natural kinds and essences, charging that their arguments are both unsound and question begging. Since Kripke and Putnam give different arguments and hold slightly different versions of the essentialist theory, we will focus exclusively on Putnam's account in what follows.

Putnam's task is to explain why the extension of natural-kind terms such as *water*, *gold* and *tiger* must include all and only those things that have the same essential properties. Let us focus on the term *water*. The traditional theory of meaning, stemming from John Locke (1632–1704) and developed by Gottlob Frege (1848–1925), maintains that extensions are determined by intensions: the intension (or meaning) of *water* is a list of properties that define what we mean by *water*. Since people were talking meaningfully about water long before the advent of modern atomic theory, the intension of *water* does not include “being H_2O .” Presumably, *water* means something like “the colorless, odorless liquid that fills lakes and rivers, falls from the sky, and is the most common solvent.” According to this traditional theory of meaning, the term *water* need not refer to the same substance, with the same microstructural properties, in all possible worlds. The most common solvent that is odorless and colorless and fills the lakes and rivers of a possible world might be something other than water—say, XYZ. So Putnam has to give us a new theory of reference in which natural kind terms, such as *water*, *gold*, and *tiger*, do not have their extensions fixed by their intensions.

PUTNAM'S NEW THEORY OF REFERENCE

Putnam's new theory of reference comes in two parts. First, he contends that a natural kind term gets its reference fixed in the actual world by means of archetypal specimens and not by mere description or intension. In this picture, we fix the reference of *water* in our world by pointing to samples of that particular liquid and (for all practical purposes) saying

"water is *this* kind of stuff, the stuff in our lakes and streams, the stuff that falls from the sky when it rains." The key here is that we fix the reference by employing the relation of *same stuff* or *same kind*—a deep relation among all those things sharing, not merely superficial characteristics, but important microstructural features. (In the present case, these important properties are taken to be the molecular properties of the compound H_2O). We thus collectively refer to all and only liquids standing in that same-kind relation to our particular sample.

Second, Putnam claims that we can then extend the reference of *water* to other possible worlds by saying (for all practical purposes) "Something is water, in any possible world, if and only if it is the same kind of stuff as this sample here." On such a view, then, nothing—in any world—can be water unless it stands in the same-kind relation to this liquid here in our glass or pitcher. That is to say, *water* refers to a single kind of stuff in every possible world: necessarily, whatever is water is H_2O . In this way, it emerges as a necessarily true law of nature that if x is water, then x contains hydrogen and oxygen.

MELLOR'S CRITICISMS OF PUTNAM'S THEORY OF REFERENCE

Mellor rejects both parts of Putnam's account. First, he denies that terms like *water* and *gold* get their reference fixed in this world by means of archetypes (rather than, say, by means of descriptions). One group of cases that Mellor thinks refutes Putnam's view are those in which there are no archetypes to point at when the term is first introduced. In an earlier paper (entitled "Natural Kinds"), he writes:

Consider elements high in the periodic table, that do not occur in nature and have never been made. We have names for them, but there may never be archetypes to constrain our use of the names. Even if specimens eventually appear, the discovery, creation or synthesis of previously unknown fundamental particles, elements and compounds can surely be *predicted*. The term 'neutrino' applied to just the same particles when it was used to predict their existence as it has applied to since their discovery. Ostensive reference (say to a bubble chamber photograph) could not have fixed its extension then; why suppose exactly the same extension is fixed that way now?⁹

Second, Mellor thinks that Putnam's analysis of the same-kind relation across possible worlds begs the question. The issue, remember, is whether the term *water* must refer to exactly the same kind of stuff, with exactly the same essential nature, in all possible worlds. Putnam's procedure relies on "important" physical properties. But there is nothing in Putnam's procedure per se that guarantees that these important properties will be exactly the same in every possible world. Suppose, for example, that all samples of water share ten important properties but that water could lack

any one of them.¹⁰ Now imagine two possible worlds, PW1 and PW2. A liquid in PW1 shares nine of these important properties with water in the actual world. Let the property missing in PW1 be P_1 . A liquid in PW2 also shares nine of these properties with water in the actual world, but in this case the missing property is P_2 . When compared with water in the actual world, the liquid in PW1 and the liquid in PW2 both qualify as water. But the two liquids would not seem to be essentially the same, since they differ in two important properties. So Putnam's procedure for fixing the reference of *water* in other possible worlds does not entail essentialism; it does not guarantee that all the things referred to by the term *water* must have exactly the same essential properties in all possible worlds.

Putnam would reject Mellor's criticism because Putnam assumes that the same-kind relation across possible worlds is an equivalence relation and all equivalence relations are transitive. If the same-kind relation is transitive, then if the liquid in PW1 is the same kind of stuff as water in the actual world and if water in the actual world is the same kind of stuff as the liquid in PW2, then the liquid in PW1 has to be the same kind of stuff as the liquid in PW2. Thus, Putnam's assumption that the same-kind relation is an equivalence relation entails that all things that belong to the same kind have to share exactly the same set of important properties.

Mellor complains that this move of Putnam's begs the question because it illicitly takes for granted the very point at issue, namely, whether water has a fixed essence. In Mellor's words, "for Putnam to claim the same-kind relation to be transitive, which he does in taking it to be an equivalence relation, is for him gratuitously to assume the essentialist conclusion he is out to prove" (857).

Mellor's criticism seems to be correct. Putnam's new theory of reference does not, by itself, guarantee the truth of essentialism. Essentialism requires that the same-kind relation be an equivalence relation and hence transitive across possible worlds. Putnam's theory of reference does not, by itself, entail that the same-kind relation is transitive because, according to that theory, things in other possible worlds count as water by comparing them with archetypes in the actual world, not by comparing them with each other. Insisting that the same-kind relation be an equivalence relation entails that all possible things that are water must have the same fundamental properties as archetypes in the actual world. But that is simply to insist on the truth of essentialism, not to give an independent argument for it. Mellor's second criticism does not prove that Putnam's theory is false, but it does leave Putnam the task of responding to Mellor's first criticism (that archetypes do not seem necessary for fixing the reference of natural kind terms). And even if Putnam were to succeed in rebutting Mellor here, we would still be a long way from the conclusion that all laws of nature are necessary truths, for Putnam's theory applies only to natural kinds (such as molecules, animals, and plants), and it is not obvious that all laws of nature are about such kinds. (For example, consider

the laws of thermodynamics and the laws of motion. None of these laws is about any specific class or category of objects having a common nature. They apply to all objects and systems *regardless* of their nature.)

7.5 | Cartwright's Antirealism about Fundamental Laws

It is noteworthy that at the end of his article, "Laws of Nature," Dretske limits himself to a conditional claim. He does not purport to have shown that there actually *are* universals and contingent relations of nomic necessitation between them. Rather, he asserts, *if* there are any laws of nature, then the universals account of them is correct. This raises the question that Cartwright addresses in "Do the Laws of Physics State the Facts?" Namely, are there laws of nature? Undoubtedly there are many generalizations in science, and many of these are called laws. But can any of these so-called laws play the role traditionally assigned to them, especially that of explanation, while at the same time being true? Do our explanatory laws truly describe how bodies actually behave? Cartwright thinks not.

Cartwright defends the following disjunction: either laws are false but can be used to explain things, or laws are true but are useless for explanation. There is, in her words, "a trade-off between factual content and explanatory power" (875). Thus, Cartwright admits that some laws—those referred to as phenomenological—can be fairly accurate descriptions of how bodies actually behave. But these phenomenological laws achieve their descriptive accuracy at the price of being highly qualified and thus highly restricted in their scope. It is the unqualified, fundamental laws of wide explanatory scope that are the target of Cartwright's antirealism. (For an extended discussion of antirealism, see chapter 9 of this volume.)

Much of Cartwright's case against fundamental laws rests on her illustration involving the laws of gravitational attraction and electrostatic attraction (and repulsion). If the gravitational law were true, then it would describe how bodies behave. In particular the law would predict the real, actual forces that act on gravitating bodies. But nearly all gravitating bodies are also electrically charged, and the smaller the body, the greater the role that charge plays in determining its behavior. Thus, taken at face value, the gravitational law seems false; it does not state correctly the actual net force acting on all gravitating bodies because the actual net force experienced by most bodies depends jointly on their mass and electrical charge.

One way of trying to retain the gravitational law as a true description of how bodies actually behave would be to limit its scope to just those bodies on which only gravitation is acting—to bodies that are not charged (or affected in any way by nongravitational forces). But that would drastically limit the number of bodies that the law describes and would render the law virtually useless for most explanatory purposes. Thus, Cartwright

concludes either the laws we use for explanation are false, or the laws are true but virtually useless for explanation: we cannot be realists about explanatory, fundamental laws.

A RESPONSE TO CARTWRIGHT'S ANTIREALIST ARGUMENT

One way of responding to Cartwright's argument for antirealism about fundamental laws is to distinguish between the actual net force acting on a body and the component force due to gravity. Indeed, this is how most of us are taught physics and mechanics. First we use individual laws to calculate the component forces acting on a body due to gravitation, electricity, tension in a spring, and so on. Then we use vector addition (the parallelogram law) to sum these forces and derive the net force acting on the body. If the net force is not zero, then the body will accelerate in accordance with Newton's second law of motion.

Cartwright denies that this response is an adequate defense of realism about fundamental laws. Her main objection is that only the net force, the actual force that determines the acceleration of a body, is real. Nature does not add component forces. Indeed, component forces are fictitious; they are not real forces at all. Why not? Because if they were real, then they would act in addition to the net force and thus give the wrong prediction about the body's motion.

Cartwright's double-counting objection to the reality of component forces seems to depend on the assumption that real forces produce actual accelerations and are measured by those accelerations. Thus, Cartwright would say that if a body were pulled in opposite directions by equal component forces and thus remained at rest, then there is no real force acting at all. If this is indeed her view, then it seems mistaken. After all, a ball on which equal and opposite forces are acting is in a different state (a state of tension) from a ball that is free from external forces. Component forces can have real effects, even if they do not produce a net acceleration.¹¹

One merit of Cartwright's paper is that it makes clear the price one must pay to be a realist about many laws in physics. Individual laws describe component forces, but component forces do not determine how bodies move (only net forces do that). So individual laws do not describe how bodies actually move. At best they describe, not actual behavior, but tendencies to behave. They specify capacities or dispositions of bodies by telling us how they would move if they were free from all other forces. And this, clearly, is a far cry from the traditional empiricist view of Hume and his followers, who would limit laws to describing what actually happens in the world. Taking laws seriously and realistically seems to require an ontology of powers and dispositions that is inconsistent with the regularity theory of laws.¹²

7.6 | Summary

Most philosophers attempting to explain what laws are adopt one of two approaches: the regularity approach or the necessitarian approach. Inspired by Hume's analysis of causation, regularity theorists (such as Ayer) insist that laws are simply descriptions or summaries of what actually happens in the world. In its simplest form, the regularity theory insists that the law that all *F*s are *G* says nothing more about the world than does the generalization that, as a matter of fact, all *F*s are *G*. This simple version of the regularity theory has a number of problems, such as the problem of accounting for those scientific laws that, like Newton's first law, have no instances. But the severest objection to the simple version of the regularity theory is that it cannot distinguish between genuine laws and accidental generalizations. In response to this difficulty, many regularity theorists impose further conditions on a true generalization before it can qualify as a law. Because these extra conditions refer to the beliefs and attitudes of scientists towards the generalization in question, this more sophisticated version of the regularity theory is often called the epistemic regularity theory. Ayer advocates one version of the epistemic regularity theory. But as Ayer acknowledges, the epistemic regularity theory makes laws inherently subjective by taking the lawlike status of a generalization to depend on whether scientists treat it as a law. Obviously, scientists cannot treat a generalization as a law if they have not yet discovered it. So the epistemic regularity theory entails that there are no unknown laws. This conclusion, that there are no unknown laws, is so counterintuitive that Dretske offers it as a sufficient reason for abandoning the entire regularity approach.

Dretske advocates a version of the necessitarian approach. Unlike older versions of this approach (which regarded laws as asserting a special relation of necessity between objects or events), Dretske's theory regards laws as relations between properties. Since properties are universals, Dretske's proposal is called the *universals theory*. One group of problems with the universals theory centers around the special relation that is supposed to hold between the universals that make up a law. Usually we think of the relations between universals as being logical relations. But if the necessitation relation were logical, then laws of nature would be necessary truths. Dretske himself rejects this, insisting that laws of nature are contingent. Against this view of the contingency of laws, some philosophers, notably Kripke and Putnam, have defended the claim that a wide class of natural laws, namely those that attribute essential properties to natural kinds, are metaphysically necessary. The case for the Kripke-Putnam doctrine rests largely on a new theory about how terms that refer to natural kinds get their meaning. Mellor, a regularity theorist, criticizes Putnam's new theory of reference on the grounds that, as a defense of essentialism,

it is question begging. If Mellor is right, then Putnam has not given a compelling reason for regarding many laws of nature as metaphysically necessary. But this still leaves necessitarians such as Dretske with the task of explaining how the special nonlogical relation characteristic of laws can hold contingently between universals. This difficulty—the problem of identification—is closely related to a second difficulty—the problem of inference. The problem of inference is the problem of explaining how, according to the necessitarian, we can validly infer that a particular thing that is *F* must also be *G* (a conclusion about a particular object) from the law that *F*-ness necessitates *G*-ness (a premise about universals). Dretske offers an analogy between laws of nature and legal codes (such as the Constitution of the United States) that define the powers and relations between the branches of government. The laws are about offices and institutions, but they imply that the people who hold those offices or serve in those institutions must do certain things. Dretske admits that the analogy is imperfect and confesses that the necessitation relation is not easy to understand or explain. Nevertheless, he thinks that the regularity theory is so flawed that the universals theory must be on the right track, despite its difficulties.

One major difference between the regularity theory and older versions of the necessitarian theory is that the necessitarian thinks that the regularity theory is too weak. The regularity theorist regards laws merely as true descriptions of how objects actually behave. The necessitarian insists that laws do more than this: they not only describe how the world is, they also assert how the world must be. But both theorists agree that a generalization must be true in order for it to be a law. Nancy Cartwright challenges this assumption. Cartwright argues that many of the laws we use to explain things are, in fact, false because they do not describe what actually happens in the world. Laws about electrostatic, magnetic, and gravitational forces, for example, do not as a rule describe how bodies actually move. Rather, they specify how bodies would move were certain ideal conditions realized. But such conditions, as a matter of fact, hardly ever obtain. Although Cartwright's argument involves some controversial claims about the nonreality of component forces, one of her conclusions seems quite plausible, namely, that many laws specify the tendencies and dispositions of bodies rather than their actual behavior. Thus, neither the regularity approach nor the older necessitarian view can be deemed adequate, since both approaches entail that laws describe how bodies do or must behave. Universals theories such as Dretske's are not inconsistent with Cartwright's position but much more needs to be said by their supporters about the nature of the two ingredients in a law—the universals (properties) and the (contingent) relation of necessitation that holds between them—and about the relation between laws and the behavior of particular objects.

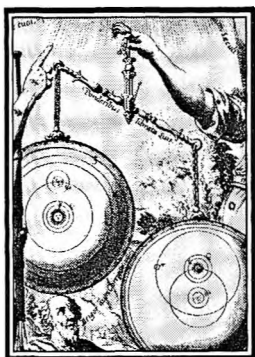
■ Notes

1. See David Hume, "Of the Idea of Necessary Connection," Section 7 of *An Enquiry Concerning Human Understanding* (1748).
2. It is important not to be misled by language here. Rather than saying (i) "It must be the case that, if Jones was elected, then he received the most votes," we might say, more idiomatically, (ii) "If Jones was elected, then he must have received the most votes." But if we add to (ii) that (iii) "Jones was indeed elected," we cannot conclude from this that "Jones *must have*—that is, *necessarily*—received the most votes." Since, in this example, (ii) is merely an equivalent way of saying (i), all that validly follows from (ii) and (iii) is that, as a matter of contingent fact, Jones received the most votes. So, too, in Dretske's example. All that follows from "It must be the case that, if x is F , then x is G " and " a is F " is that, as a matter of contingent fact, a is G , not that a must be G .
3. See the opening propositions of Ludwig Wittgenstein, *Tractatus Logico-Philosophicus*, trans. D. F. Pears and B. F. McGuinness (London: Routledge and Kegan Paul, 1961), p. 7.
4. For a lively attack by a fellow necessitarian on the contention (by Dretske, Armstrong, and others) that nomic necessitation between the universals of a law of nature is always contingent, see Martin Tweedale, "Universals and Laws of Nature," *Philosophical Topics* 13 (1982): 25–44.
5. Part of the difficulty with Hume's conceivability test is its vagueness. For it is one thing to conceive that something is water without also conceiving that it contains oxygen. Anyone who is ignorant of chemistry could do this. But can I conceive that something is water and that it does not contain oxygen? The problem is the unclarity surrounding the notion that I could properly be said to be conceiving of water—real, actual water—when I imagine a liquid that looks like water but does not contain oxygen.
6. See Saul A. Kripke, *Naming and Necessity* (Cambridge, Mass.: Harvard University Press, 1972).
7. For a good introduction see Kripke, *Naming and Necessity*, and Stephen P. Schwartz, ed., *Naming, Necessity, and Natural Kinds* (Ithaca, N.Y.: Cornell University Press, 1977).
8. A word of caution: Kripke and Putnam recognize that our best scientific theories might be mistaken. So, strictly speaking, their claim is conditional: if modern science is right about the nature of gold, then the law that gold has an atomic number of 79 is a necessary truth.
9. D. H. Mellor, "Natural Kinds," *British Journal for the Philosophy of Science* 28 (1977): 306.
10. The example and argument are from Mellor's "Natural Kinds." If Mellor's example seems implausible, consider biological species. Surely, not all tigers have exactly the same set of genes; but, presumably, tigers share a sufficiently large number of genes (or very similar genes) to be members of the same species. Similarly, we know that the stuff we usually call *water* is in fact a mixture of three

compounds—hydrogen oxide, deuterium oxide, and tritium oxide—and although hydrogen oxide is by far the most common of the three in the actual world, it is logically possible that the ratio might vary. Moreover, it is at least imaginable that advances in elementary particle physics might reveal that not even all molecules of hydrogen oxide have exactly the same internal structure.

11. For more on Cartwright's double-counting objection to the reality of component forces, see Lewis Creary, "Causal Explanation and the Reality of Natural Component Forces," *Pacific Philosophical Quarterly* 62 (1981): 148–57; A. David Kline and Carl A. Matheson, "How the Laws of Physics Don't Even Fib," in *PSA 1986*, ed. A. Fine and P. Machamer (East Lansing, Mich.: Philosophy of Science Association, 1986), 1: 33–41. Cartwright replies to Creary in Nancy Cartwright, *How the Laws of Physics Lie* (Oxford: Clarendon Press, 1983), 62–67.

12. Cartwright develops such a view, which she traces back to John Stuart Mill, in Nancy Cartwright, *Nature's Capacities and Their Measurement* (Oxford: Clarendon Press, 1989). There has been a debate about whether Cartwright's realism about capacities (which includes such things as dispositions, powers, and tendencies) is consistent with her antirealism about fundamental laws (which she continues to maintain). See Alan Chalmers, "So the Laws of Physics Needn't Lie," *Australasian Journal of Philosophy* 71 (1993): 196–205; Steve Clarke, "The Lies Remain the Same: A Reply to Chalmers," *Australasian Journal of Philosophy* 73 (1995): 152–55, and Alan Chalmers, "Cartwright on Fundamental Laws: A Response to Clarke," *Australasian Journal of Philosophy* 74 (1996): 150–52.



8 | Intertheoretic Reduction

INTRODUCTION

Reduction is a perennial theme in philosophy, where it crops up regularly in fields as diverse as epistemology, metaphysics, the philosophy of mind, and the philosophy of mathematics. But reduction has a special significance for the philosophy of science because several philosophers (notably Ernest Nagel) have claimed that reduction holds the key to understanding an important kind of progress in science.

According to Nagel and others, when scientists discover a new theory that is able to reduce or absorb one or more already well-established and accepted theories, then we have a paradigm instance of objective, cumulative progress. This account of scientific progress is sometimes called the *Russian doll model*, for just as each doll nests inside the next larger one, so, too, do later theories include within them all the earlier theories that they have absorbed. In this way, science is supposed to display a steady increase in empirical content and explanatory power. In the nineteenth century, for example, classical thermodynamics, with its distinctive concepts of heat, entropy, and temperature, became integrated into—was absorbed by—statistical mechanics, a new and more powerful branch of physics applying probability theory to molecular motion. In a similar way, twentieth-century molecular biology has managed to incorporate most of classical genetics, largely through the Watson-Crick discovery of the structure of DNA.

In “Issues in the Logic of Reductive Explanations,” Nagel distinguishes between two sorts of intertheoretic reduction—homogeneous and inhomogeneous—and specifies the formal conditions that each must satisfy. Nagel views reduction as a type of explanation and accepts Hempel’s deductive-nomological model of scientific explanation. (See chapter 6 for details of Hempel’s model.) Thus, Nagel requires that in both homoge-

neous and inhomogeneous reduction the reduced theory be logically derivable from the reducing theory. In inhomogeneous reductions, there are terms in the reduced theory that are absent from the reducing theory. Consequently, these reductions also need *rules of correspondence* or *bridge laws* to make possible the derivation that Nagel requires. Another important constraint on reduction in Nagel's model is that the meanings of theoretical terms not change when reduction occurs.

Nagel's formal model of reduction has not gone unchallenged. Principal among its critics has been Paul Feyerabend, whose attack on Nagel is conveyed in his paper, "How to Be a Good Empiricist—A Plea for Tolerance in Matters Epistemological." Much of Feyerabend's case against Nagel rests on Feyerabend's controversial holist thesis that the meaning of any theoretical term always depends on the theory in which it occurs.

Many philosophers, even those who reject Feyerabend's views about meaning, agree that Nagel's account of reduction is flawed. But they think that the project of trying to understand the nature of intertheoretic reduction is valuable and should be pursued. Typical among these philosophers is Thomas Nickles. In "Two Concepts of Intertheoretic Reduction," Nickles offers an analysis of two different varieties of reductive relation that are intended to avoid Feyerabend's objections while at the same time reflecting what actually goes on in science.

Finally, in the last piece in the chapter, "1953 and All That: A Tale of Two Sciences," Philip Kitcher casts a critical eye on the claim, oft-repeated, that classical genetics has been reduced to molecular biology. Kitcher argues that modern biology fails to match any of the available models of reduction. Instead, he advocates a new way of thinking about scientific theories in order to understand the relations between a theory and its successor. Echoing some of the views of Thomas Kuhn, Kitcher recommends that theories should be understood in terms of practices, problems, and patterns of reasoning. Only then, he thinks, can we do justice to the complexity of the relationship between theories in modern biology.

ERNEST NAGEL

Issues in the Logic of Reductive Explanations

A recurrent theme in the long history of philosophical reflection on science is the contrast—voiced in many ways by poets and scientists as well as philosophers—between the characteristics commonly attributed to things on the basis of everyday encounters with them, and the accounts of those things given by scientific theories that formulate some ostensibly pervasive executive order of nature. This was voiced as early as Democritus, when he declared that while things are customarily said to be sweet or bitter, warm or cold, of one color rather than another, in truth there are only the atoms and the void. The same contrast was implicit in Galileo's distinction, widely accepted by subsequent thinkers, between the primary and secondary qualities of bodies. It was dramatically stated by Sir Arthur Eddington in terms of currently held ideas in physics, when he asked which of the two tables at which he was seated was "really there"—the solid, substantial table of familiar experience, or the insubstantial scientific table which is composed of speeding electric charges and is therefore mostly "emptiness."^{*}

FROM *Teleology Revisited* (New York: Columbia University Press, 1974), 95–113.

^{*} See the introduction to Eddington's Gifford Lectures, published as *The Nature of the Physical World* (New York: Macmillan, 1928). Eddington (1882–1944) was Plumian Professor of Astronomy at Cambridge and made important contributions to relativity theory and astrophysics. In his popular writings, Eddington argued for the idealist conclusion that "the stuff of the world is mind-stuff" partly on the grounds that modern physics (relativity and quantum theory) has refuted materialism. For example, Eddington claimed that modern science had shown that the first of his two tables—the solid, substantial table of everyday life—is a fiction and that what really exists is the scientific table, which is mostly empty space occupied by electrons and other insubstantial particles. Beginning with Lizzie Susan Stebbing (1885–1943), philosophers have been critical of Eddington's attempt to derive idealist conclusions from the success of reductionist theories in modern physics.

Formulations of the contrast vary and have different overtones. In some cases, as in the examples I have cited, the contrast is associated with a distinction between what is allegedly only "appearance" and what is "reality"; and there have been thinkers who have denied that so-called "common-sense" deals with ultimate reality, just as there have been thinkers who have denied that the statements of theoretical science do so. However, a wholesale distinction between appearance and reality has never been clearly drawn, especially since these terms have been so frequently used to single out matters that happen to be regarded as important or valuable; nor have the historical controversies over what is to count as real and what as appearance thrown much light on how scientific theories are related to the familiar materials that are usually the points of departure for scientific inquiry. In any case, the contrast between the more familiar and manifest traits of things and those which scientific theory attributes to them need not be, and often is not, associated with the distinction between the real and the apparent; and in point of fact, most current philosophies of science, which in one way or another occupy themselves with this contrast, make little if any use of that distinction in their analyses.

But despite important differences in the ways in which the contrast has been formulated, I believe they share a common feature and can be construed as being addressed to a common problem. They express the recognition that certain relations of dependence between one set of distinctive traits of a given subject matter are allegedly explained by, and in some sense "reduced" to, assumptions concerning more inclusive relations of dependence between traits or processes not distinctive of (or unique to) that subject matter. They implicitly raise the question of what, in fact, is the logical structure of such reductive explanations—whether they differ from other sorts of scientific explanation, what is achieved by reductions, and under what conditions they are feasible. These questions are important for the understanding of modern science, for its development is marked by strong reductive tendencies, some of whose outstanding achievements are often counted as examples of reduction. For example, as a consequence of this reductive process, the theory of heat is commonly said to be but a branch of Newtonian mechanics, physical optics a branch of electromagnetic theory, and chemical laws a branch of quantum mechanics. Moreover, many biological processes have been given physico-chemical explanations, and there is a continuing debate as to the possibility of giving such explanations for the entire domain of biological phenomena. There have been repeated though still unsuccessful attempts

See L. S. Stebbing, *Philosophy and the Physicists* (London: Methuen, 1937). For some related criticisms, see "NOA's Ark—Fine for Realism," (in chapter 9), where Alan Musgrave diagnoses Eddington's mistake as the error of supposing that to explain something (such as the solidity of table) is to explain it away (that is, to show that it does not really exist).

to exhibit various patterns of men's social behavior as examples of psychological laws.

It is with some of the issues that have emerged in proposed analyses of reductive explanations that this paper is concerned. I will first set out in broad outlines what I believe is the general structure of such explanations; then examine some difficulties that have recently been raised against this account. . . .

■ | 1

Although the term "reduction" has come to be widely used in philosophical discussions of science, it has no standard definition. It is therefore not surprising that the term encompasses several sorts of things which need to be distinguished. But before I do this, a brief terminological excursion is desirable. Scientists and philosophers often talk of deducing or inferring one phenomenon from another (e.g., of deducing a planet's orbital motion), of explaining events or their concatenations (e.g., of explaining the occurrence of rainbows), and of reducing certain processes, things, or their properties to others (e.g., of reducing the process of heat conduction to molecular motions). However, these locutions are elliptical, and sometimes lead to misconceptions and confusions. For strictly speaking, it is not phenomena which are deduced from other phenomena, but rather *statements* about phenomena from other statements. This is obvious if we remind ourselves that a given phenomenon can be subsumed under a variety of distinct descriptions, and that phenomena make no assertions or claims. Consequently, until the traits or relations of a phenomenon which are to be discussed are indicated, and predications about them are formulated, it is literally impossible to make any deductions from them. The same holds true for the locutions of explaining or reducing phenomena. I will therefore avoid these elliptic modes of speech hereafter, and talk instead of deducing, explaining, or reducing statements about some subject matter.

Whatever else may be said about reductions in science, it is safe to say that they are commonly taken to be explanations, and I will so regard them. In consequence, I will assume that, like scientific explanations in general, every reduction can be construed as a series of statements, one of which is the conclusion (or the statement which is being reduced), while the others are the premises or reducing statements. Accordingly, reductions can be conveniently classified into two major types: homogeneous reductions, in which all of the "descriptive" or specific subject matter terms in the conclusion are either present in the premises also or can be explicitly defined using only terms that are present; and inhomogeneous reductions, in which at least one descriptive term in the conclusion

neither occurs in the premises nor is definable by those that do occur in them. I will now characterize in a general way what I believe to be the main components and the logical structure of these two types of reduction, but will also state and comment upon some of the issues that have been raised by this account of reduction.

A frequently cited example of homogeneous reduction is the explanation by Newtonian mechanics and gravitational theory of various special laws concerning the motions of bodies, including Galileo's law for freely falling bodies near the earth's surface and the Keplerian laws of planetary motion. The explanation is homogeneous, because on the face of it at any rate, the terms occurring in these laws (e.g., distance, time, and acceleration) are also found in the Newtonian theory. Moreover, the explanation is commonly felt to be a reduction of those laws, in part because these laws deal with the motions of bodies in restricted regions of space which had traditionally been regarded as essentially dissimilar from motions elsewhere (e.g., terrestrial as contrasted with celestial motions), while Newtonian theory ignores this traditional classification of spatial regions and incorporates the laws into a unified system. In any event, the reduced statements in this and other standard examples of homogeneous reduction are commonly held to be deduced logically from the reducing premises. In consequence, if the examples can be taken as typical, the formal structure of homogenous reductions is in general that of deductive explanations. Accordingly, if reductions of this type are indeed deductions from theories whose range of application is far more comprehensive and diversified than that of the conclusions derived from them, homogeneous reductions appear to be entirely unproblematic, and to be simply dramatic illustrations of the well understood procedure of deriving theorems from assumed axioms.

However, the assumption that homogeneous reductions are deductive explanations has been recently challenged by a number of thinkers, on the ground that even in the stock illustrations of such reductions the reduced statements do not in general follow from the explanatory premises. For example, while Galileo's law asserts that the acceleration of a freely falling body near the earth's surface is constant, Newtonian theory entails that the acceleration is not constant, but varies with the distance of the falling body from the earth's center of mass. Accordingly, even though the Newtonian conclusion may be "experimentally indistinguishable" from Galileo's law, the latter is in fact "inconsistent" with Newtonian theory. Since it is this theory rather than Galileo's law that was accepted as sound, Galileo's law was therefore *replaced* by a different law for freely falling bodies, namely the law derived from the Newtonian assumptions. A similar outcome holds for Kepler's third planetary law. The general thesis has therefore been advanced that homogeneous reductions do not consist in the deduction or explanation of laws, but in the total *replacement* of incorrect assumptions by radically new ones which are believed to be more

correct and precise than those they replace. This thesis raises far-reaching issues, and I will examine some of them presently. But for the moment I will confine my comments on it to questions bearing directly on homogeneous reductions.

a) It is undoubtedly the case that the laws derivable from Newtonian theory do not coincide exactly with some of the previously entertained hypotheses about the motions of bodies, though in other cases there may be such coincidence. This is to be expected. For it is a widely recognized function of comprehensive theories (such as the Newtonian one) to specify the conditions under which antecedently established regularities hold, and to indicate, in the light of those conditions, the modifications that may have to be made in the initial hypotheses, especially if the range of application of the hypotheses is enlarged. Nevertheless, the initial hypotheses may be reasonably close approximations to the consequences entailed by the comprehensive theory, as is indeed the case with Galileo's law as well as with Kepler's third law. (Incidentally, when Newtonian theory is applied to the motions of just two bodies, the first and second Keplerian laws agree fully with the Newtonian conclusions). But if this is so, it is correct to say that in homogeneous reductions the reduced laws are either derivable from the explanatory premises, or are good approximations to the laws derivable from the latter.

b) Moreover, it is pertinent to note that in actual scientific practice, the derivation of laws from theories usually involves simplifications and approximations of various kinds, so that even the laws which are allegedly entailed by a theory are in general only approximations to what is strictly entailed by it. For example, in deriving the law for the period of a simple pendulum, the following approximative assumptions are made; the weight of the pendulum is taken to be concentrated in the suspended bob; the gravitational force acting on the bob is assumed to be constant, despite variations in the distance of the bob from the earth's center during the pendulum's oscillation; and since the angle through which the pendulum may oscillate is stipulated to be small, the magnitude of the angle is equated to the sine of the angle. The familiar law that the period of a pendulum is proportional to the square root of its length divided by the constant of acceleration is therefore derivable from Newtonian theory only if these various approximations are taken for granted.

More generally, though no statistical data are available to support the claim, there are relatively few deductions from the mathematically formulated theories of modern physics in which analogous approximations are not made, so that many if not all the laws commonly said by scientists to be deducible from some theory are not strictly entailed by it. It would nevertheless be an exaggeration to assert that in consequence scientists are fundamentally mistaken in claiming to have made such deductions. It is obviously important to note the assumptions, including those concerning approximations, under which the deduction of a law is made. But it does

not follow that given those assumptions a purported law cannot count as a consequence of some theory. Nor does it follow that if in a proposed homogeneous reduction the allegedly reduced law is only an approximation to what is actually entailed by the reducing theory when *no* approximative assumptions are added to the latter, the law has not been reduced but is being replaced by a radically different one.

c) Something must also be said about those cases of homogeneous reduction in which the law actually derivable from the reducing theory makes use of concepts not employed in the law to be reduced. Thus, while according to Kepler's third (or harmonic) law, the squares of the periods of the planets are to each other as the cubes of their mean distances from the sun, the Newtonian conclusion is that this ratio is not constant for all the planets but varies with their *masses*. But the notion of mass was introduced into mechanics by Newton, and does not appear in the Keplerian law; and although the masses of the planets are small in comparison with the mass of the sun, and the Keplerian harmonic law is therefore a close approximation to the Newtonian one, the two cannot be equated. Nevertheless, while the two are not equivalent, neither are they radically disparate in content or meaning. On the contrary, the Newtonian law identifies a causal factor in the motions of the planets which was unknown to Kepler.

■ | 2

I must now turn to the second major type of reductive explanations. Inhomogeneous reductions, perhaps more frequently than homogeneous ones, have occasioned vigorous controversy among scientists as well as philosophers concerning the cognitive status, interpretation, and function of scientific theories; the relations between the various theoretical entities postulated by these theories, and the familiar things of common experience; and the valid scope of different modes of scientific analysis. These issues are interconnected, and impinge in one way or another upon questions about the general structure of inhomogeneous reductions. Since none of the proposed answers to these issues has gained universal assent, the nature of such reductions is still under continuing debate.

Although there are many examples of inhomogeneous reductions in the history of science, they vary in the degree of completeness with which the reduction has been effected. In some instances, all the assumed laws in one branch of inquiry are apparently explained in terms of a theory initially developed for a different class of phenomena; in others, the reduction has been only partial, though the hope of completely reducing the totality of laws in a given area of inquiry to some allegedly "basic" theory may continue to inspire research. Among the most frequently cited

illustrations of such relatively complete inhomogeneous reductions are the explanation of thermal laws by the kinetic theory of matter, the reduction of physical optics to electromagnetic theory, and the explanation (at least in principle) of chemical laws in terms of quantum theory. On the other hand, while some processes occurring in living organisms can now be understood in terms of physicochemical theory, the reducibility of all biological laws in a similar manner is still a much disputed question.

In any case, the logical structure of inhomogeneous reductive explanations is far less clear and is more difficult to analyze than is the case with homogeneous reductions. The difficulty stems largely from the circumstance that in the former there are (by definition) terms or concepts in the reduced laws (e.g., the notion of heat in thermodynamics, the term "light wave" in optics, or the concept of valence in chemistry) which are absent from the reducing theories. Accordingly, if the overall structure of the explanation of laws is taken to be that of a deductive argument, it seems impossible to construe inhomogeneous reductions as involving essentially little more than the logical derivation of the reduced laws (even when qualifications about the approximative character of the latter are made) from their explanatory premises. If inhomogeneous reductions are to be subsumed under the general pattern of scientific explanations, it is clear that additional assumptions must be introduced as to how the concepts characteristically employed in the reduced laws, but not present in the reducing theory, are connected with the concepts that do occur in the latter.

Three broad types of proposals for the structure of inhomogeneous reductions can be found in the recent literature of the philosophy of science. The first, which for convenience will be called the "instrumentalist" analysis, is usually advocated by thinkers who deny a cognitive status to scientific laws or theories, regarding them as neither true nor false but as rules (or "inference tickets") for inferring so-called "observation statements" (statements about particular events or occurrences capable of being "observed" in some not precisely defined sense) from other such statements. According to this view, for example, the kinetic theory of gases is not construed as an account of the composition of gases. It is taken to be a complex set of rules for predicting, among other things, what the pressure of a given volume of gas will be if its temperature is kept constant but its volume is diminished. However, the scope of application of a given law or theory may be markedly more limited than the scope of another. The claim that a theory *T* (e.g., the corpus of rules known as thermodynamics) is reduced to another theory *T'* (e.g., the kinetic theory of gases) would therefore be interpreted as saying that all the observation statements which can be derived from given data with the help of *T* can also be derived with the help of *T'*, but not conversely. Accordingly, the question to which this account of inhomogeneous reduction is addressed is not the ostensibly asserted content of the theories involved in reduction, but the

comparative ranges of observable phenomena to which two theories are applicable.

Although this proposed analysis calls attention to an important function of theories and provides a rationale for the reduction of theories, its adequacy depends on the plausibility of uniformly interpreting general statements in science as rules of inference. Many scientists certainly do not subscribe to such an interpretation, for they frequently talk of laws as true and as providing at least an approximately correct account of various relations of dependence among things. In particular, this interpretation precludes the explanation of macro-states of objects in terms of unobservable microprocesses postulated by a theory. Moreover, the proposal is incomplete in a number of ways: it has nothing to say about how theoretical terms in laws (e.g., "electron" or even "atom") may be used in connection with matters of observation, or just how theories employing such notions operate as rules of inference; and it ignores the question of how, if at all, the concepts of a reduced theory are related to those of the reducing one, or in what way statements about a variety of observable things may fall within the scope of both theories. In consequence, even if the proposed analysis is adequate for a limited class of reductive explanations, it does not do justice to important features characterizing many others.

The second proposed analysis of inhomogeneous reductions (hereafter to be referred to—perhaps misleadingly—as the "correspondence" proposal) is also based on several assumptions. One of them is that the terms occurring in the conclusion but not in the premises of a reduction have "meanings" (i.e., uses and applications) which are determined by the procedures and definitions of the discipline to which reduced laws initially belong, and can be understood without reference to the ideas involved in the theories to which the laws have been reduced. For example, the term "entropy" as used in thermodynamics is defined independently of the notions characterizing statistical mechanics. Furthermore, the assumption is made that many subject matter terms common to both the reduced and reducing theories—in particular, the so-called observation terms employed by both of them to record the outcome of observation and experiment—are defined by procedures which can be specified independently of these theories and, in consequence, have "meanings" that are neutral with respect to the differences between the theories. For example, the terms "pressure" and "volume change" which occur in both thermodynamics and the kinetic theory of gases are used in the two theories in essentially the same sense. It is important to note, however, that this assumption is compatible with the view that even observation terms are "theory impregnated," so that such terms are not simply labels for "bare sense-data," but predicate characteristics that are not immediately manifest and are defined on the basis of various theoretical commitments. For example, if the expression "having a diameter of five inches" is counted as an observation predicate, its application to a given object implicitly involves commitment

to some theory of spatial measurement as well as to some laws concerning the instrument used in making the measurement. Accordingly, the point of the assumption is not that there are subject-matter terms whose meanings or uses are independent of *all* theories, but rather that every such term has a meaning which is fixed by *some* theory but independent of others. A third assumption underlying the correspondence analysis of inhomogeneous reductions is that, like homogeneous reduction, and with similar qualifications referring to approximations, they embody the pattern of deductive explanations.

In view of these assumptions, it is clear that if a law (or theory) *T* is to be reduced to a theory *T'* not containing terms occurring in *T*, *T'* must be supplemented by what have been called "rules of correspondence" or "bridge laws," which establish *connections* between the distinctive terms of *T* and certain terms (or combinations of terms) in *T'*. For example, since the second law of thermodynamics talks of the transfer of heat, this law cannot be deduced from classical mechanics, which does not contain the term "heat," unless the term is connected in some way with some complex of terms in mechanics. The statement of such a connection is a correspondence rule. However, because of the first of the above three assumptions, a correspondence rule cannot be construed as an explicit definition of a term distinctive of *T*, which would permit the elimination of the term on *purely logical grounds* in favor of the terms in *T'*. Thus, the notion of entropy as defined in thermodynamics can be understood and used without any reference to notions employed in theories about the microstructure of matter; and no amount of logical analysis of the concept of entropy can show the concept to be constituted out of the ideas employed in, say, statistical mechanics. If this is indeed the case (as I believe it is), then the theory *T* is not derivable from (and hence not reducible to) the theory *T'*, although *T* may be derivable from *T'* when the latter is conjoined with an appropriate set of bridge laws.

What then is the status of the correspondence rules required for inhomogeneous reduction? Different articulations of the theories involved in a reduction, as well as different stages in the development of inquiry into the subject matter of the theories, may require different answers; but I will ignore these complications. In general, however, correspondence rules formulate *empirical hypotheses*—hypotheses which state certain relations of dependence between things mentioned in the reduced and reducing theories. The hypotheses are, for the most part, not testable by confronting them with observed instances of the relations they postulate. They are nevertheless not arbitrary stipulations, and as with many other scientific laws their factual validity must be assessed by comparing various consequences entailed by the system of hypotheses to which they belong with the outcome of controlled observations. However, bridge laws have various forms; and while no exhaustive classification of their structure is available, two sorts of bridge laws must be briefly described.

a) A term in a reduced law may be a predicate which refers to some distinctive *attribute* or characteristic of things (such as the property of having a certain temperature or of being red) that is not denoted by any of the predicates of the reducing theory. In this case the bridge law may specify the conditions, formulated in terms of the ideas and assumptions of the reducing theory, under which the attribute occurs. For example, the kinetic theory of gases formulates its laws in terms of such notions as molecule, mass, and velocity, but does not employ the thermodynamical notion of temperature. However, a familiar bridge law states that a gas has a certain temperature when the mean kinetic energy of its molecules has a certain magnitude. In some cases, bridge laws of the sort being considered may specify conditions for the occurrence of an attribute which are necessary as well as sufficient; in other cases the conditions specified may be sufficient without being necessary; and in still other cases, the conditions stated may only be necessary. In the latter case, however, laws involving the attribute will, in general, not be deducible from the proposed reducing theory. (Thus, though some of the necessary conditions for objects having colors can be stated in terms of ideas belonging to physical optics in its current form, the physiological equipment of organisms which must also be present for the occurrence of colors cannot be described in terms of those ideas. Accordingly, if there are any laws about color relations, they are not reducible to physical optics.)

In any case, such bridge laws are empirical hypotheses concerning the *extensions* of the predicates mentioned in these correspondence rules—that is, concerning the classes of individual things or processes designated by those predicates. An attribute of things connoted by a predicate in a reduced law may indeed be quite different from the attribute connoted by the predicates of the reducing theory; but the class of things possessing the former attribute may nevertheless coincide with (or be included in) the class of things which possess the property specified by a complex predicate in the reducing theory. For example, the statement that a liquid is viscous is not equivalent in meaning to the statement that there are certain frictional forces between the layers of molecules making up the liquid. But if the bridge laws connecting the macro-properties and the microstructure of liquids is correct, the extension of the predicate “viscous” coincides with (or is included in) the class of individual systems with that microstructure.

b) Let me now say something about a second sort of correspondence rule. Although much scientific inquiry is directed toward discovering the determining conditions under which various traits of things occur, some of its important achievements consist in showing that things and processes initially assumed to be distinct are in fact the same. A familiar example of such an achievement is the discovery that the Morning Star and the Evening Star are not different celestial objects but are identical. Similarly, although the term “molecule” designates one class of particles and the

term "atom" designates another class, molecules are structures of atoms, and in particular a water molecule is an organization of hydrogen and oxygen atoms denoted by the formula " H_2O "; and accordingly, the extension of the predicate "water molecule" is the same as the class of things designated by the formula. Correspondence rules of the second sort establish analogous identifications between classes of individuals or "entities" (such as spatiotemporal objects, processes, and forces) designated by different predicates. An oft cited example of such rules is a bridge law involved in the reduction of physical optics to electromagnetic theory. Thus, prior to Maxwell, physicists postulated the existence of certain physical propagations designated as "light waves," while electromagnetic theory was developed on the assumption that there are electromagnetic waves. An essential step in the reduction of optics to electrodynamics was the introduction by Maxwell of the hypothesis (or bridge law) that these are not two *different* processes but a *single* one, even though electromagnetic waves are not always manifested as visible light. Analogous bridge laws are assumed when a flash of lightning is said to be a surge of electrically charged particles, or when the evaporation of a liquid is asserted to be the escape of molecules from its surface; and while the full details for formulating a similar bridge law are not yet available, the hope of discovering them underlies the claim that a biological cell is a complex organization of physicochemical particles.

Correspondence rules of the second kind thus differ from rules of the first, in that unlike the latter (which state conditions, often in terms of the ideas of a micro-theory, for the occurrence of traits characterizing various things, often macroscopic ones), they assert that certain logically non-equivalent expressions describe identical entities. Although both sorts of rules have a common function in reduction and both are in general empirical assumptions, failure to distinguish between them is perhaps one reason for the persistence of the mistaken belief that reductive explanations establish the "unreality" of those distinctive traits of things mentioned in reduced laws.

■ | 3

This account of inhomogeneous reduction has been challenged by a number of recent writers who have advanced an alternate theory which rejects the main assumptions of both the instrumentalist and the correspondence analyses, and which I will call the "replacement" view. Since I believe the correspondence account to be essentially correct, I shall examine the fundamental contention of the replacement thesis, as presented by Professor Paul Feyerabend, one of its most vigorous proponents.

Feyerabend's views on reduction rest upon the central (and on the

face of it, sound) assumption that "the meaning of every term we use depends upon the theoretical context in which it occurs."¹ This claim is made not only for "theoretical" terms like "neutrino" or "entropy" in explicitly formulated scientific theories, but also for expressions like "red" or "table" used to describe matters of common observation (i.e., for observation terms). Indeed, Feyerabend uses the word "theory" in a broad sense, to include such things as myths and political ideas.² He says explicitly that "even everyday languages, like languages of highly theoretical systems, have been introduced in order to give expression to some theory or point of view, and they therefore contain a well-developed and sometimes very abstract ontology."³ "The description of every single fact," he declares, is "dependent on *some* theory."⁴ He further maintains that "theories are meaningful independent of observations; observational statements are not meaningful unless they have been connected with theories."⁵ There is, therefore, no "observation core," even in statements of perception, that is independent of theoretical interpretation,⁶ so that strictly speaking each theory determines its own distinctive set of observation statements. And while he allows that two "low level" theories which fall within the conceptual framework of a comprehensive "background theory" may have a common interpretation for their observation statements, two "high level" theories concerning the nature of the basic elements of the universe "may not share a single observational statement."⁷ It is therefore allegedly an error to suppose that the empirical adequacy of a theory can be tested by appeal to observation statements whose meanings are independent of the theory, and which are neutral as between that theory and some alternative competing theory. "The methodological unit to which we must refer when discussing questions of test and empirical context, is constituted by a *whole set of partly overlapping, factually adequate, but mutually inconsistent theories.*"⁸

Moreover, a change in a theory is accompanied by a change in the meanings of all its terms, so that theories constructed on "mutually inconsistent principles" are in fact "incommensurable."⁹ Thus, if *T* is classical celestial mechanics, and *T'* is the general theory of relativity, "the meanings of all descriptive terms of the two theories, primitive as well as defined terms, will be different," the theories are incommensurable, and "not a single descriptive term of *T* can be incorporated into *T'*."¹⁰ In consequence, Feyerabend believes the correspondence account of inhomogeneous reduction is basically mistaken in supposing that allegedly reduced laws or theories can be derived from the reducing theory with the help of appropriate bridge laws:

What happens . . . when transition is made from a theory *T'* to a wider theory *T* (which, we shall assume, is capable of covering all the phenomena that have been covered by *T'*) is something much more radical than incorporation of the *unchanged* theory *T'* (unchanged, that is, with respect to the meanings

of its main descriptive terms as well as to the meanings of the terms of its observation language) into the context of *T*. What does happen is, rather, a *complete replacement* of the ontology (and perhaps even of the formalism) of *T'* by the ontology (and the formalism) of *T* and a corresponding change of the meanings of the descriptive elements of the formalism of *T'* (provided these elements and this formalism are still used). This replacement affects not only the theoretical terms of *T'* but also at least some of the observational terms which occurred in its test statements. . . . In short: introducing a new theory involves changes of outlook both with respect to the observable and with respect to the unobservable features of the world, and corresponding changes in the meaning of even the most "fundamental" terms of the language employed.^{11*}

Accordingly, if these various claims are warranted, there is not and cannot be any such thing as the reduction of laws or theories; and the examples often cited as instances of reduction are in fact instances of something else: the exclusion of previously accepted hypotheses from the corpus of alleged scientific knowledge, and the substitution for them of incommensurably different ones.

But are these claims warranted? I do not believe they are. Feyerabend is patently sound in maintaining that no single statement or any of its constituent terms has a meaning in isolation, or independently of various rules or conventions governing its use. He is no less sound in noting that the meaning of a word may change when its range of application is altered. However, these familiar truisms do not support the major conclusion he draws from them. The presentation of his thesis suffers from a number of unclaritys (such as what is to count as a change in a theory, or what are the criteria for changes in meaning), which cloud the precise import of some of his assertions. I shall, however, ignore these unclaritys here¹² and will comment briefly only on two difficulties in Feyerabend's argument.

a) It is a major task of scientific inquiry to assess the adequacy of proposed laws to the "facts" of a subject matter as established by observation or experiment, and to ascertain whether the conclusions reached are consistent with one another. However, if two proposed theories for some given range of phenomena share no term with the same meaning in each of them, so that the theories have completely different meanings (as Feyerabend believes is commonly the case), it is not evident in what sense two such theories can be said to be either compatible or inconsistent with one another: for relations of logical opposition obtain only between statements whose terms have common meanings. Moreover, it is also difficult to understand how, if the content of observation statements is determined by the theory which is being tested (as Feyerabend maintains),

* Unlike Nagel, Feyerabend uses *T'* to denote the reduced theory and *T* to denote the reducing theory.

those statements can serve as a basis for deciding between the theory and some alternative to it. For according to his analysis those observation statements will automatically corroborate the theory that happens to be used to interpret observational data, but will be simply irrelevant in assessing the empirical validity of an alternative theory. Theories thus appear to be self-certifying, and to be beyond the reach of criticism based on considerations that do not presuppose them. This outcome is reminiscent of Karl Mannheim's claim that truth in social matters is "historically relative": there are no universally valid analyses of social phenomena, since every such analysis is made within some distinctive social perspective which determines the meaning as well as the validity of what is said to be observed, so that those who do not share the same perspective can neither reach common conclusions about human affairs, nor significantly criticize each others' findings.*

Feyerabend attempts to escape from such skeptical relativism by involving what he calls the "pragmatic theory of observation." In this theory, it is still the case that the meaning of an observation statement varies with the theory used to interpret observations. However, it is possible to describe the observational and predictive statements an investigator utters as *responses* to the situations which "prompt" the utterances, and to compare the order of these responses with the order of the physical situations that prompt them, so as to ascertain the agreements or disagreements between the two orders.¹³ But if this account of the role of observation statements in testing theories is to outflank the relativism Feyerabend wants to avoid, the *secondary* statements (they are clearly observation statements) about the responses (or primary observation statements) of investigators cannot have meanings dependent on the theory being tested, and must be invariant to alternative theories. However, if secondary statements have this sort of neutrality, it is not evident why only such observation statements can have this privileged status.

b) Feyerabend has difficulties in providing a firm observational basis for objectively evaluating the empirical worth of proposed hypotheses. The difficulties stem from what I believe is his exaggerated view that the meaning of every term occurring in a theory or in its observation statements is wholly and uniquely determined by that theory, so that its meaning is

* See Karl Mannheim, *Ideologie und utopie* (Bonn: F. Cohen, 1929); *Ideology and Utopia*, trans. L. Wirth and E. Shils (New York: Harcourt Brace and Co., 1936). Mannheim (1893–1947) was the founding father of the sociology of knowledge, the discipline concerned with the social causes of beliefs as opposed to the reasons that people might have for them. As Nagel notes, Mannheim denied the possibility of objective "scientific" knowledge of social phenomena on the grounds that all judgments about society are relative to the historical period in which they are made. For criticism of a modern version of the sociology of scientific knowledge, see the discussion of the strong programme in Laudan's "Demystifying Underdetermination" in chapter 3 and in the accompanying commentary.

radically changed when the theory is modified. For theories are not quite the monolithic structures he takes them to be—their component assumptions are, in general, logically independent of one another, and their terms have varying degrees of dependence on the theories into which they enter. Some terms may indeed be so deeply embedded in the totality of assumptions constituting a particular theory that they can be understood only within the framework of the theory: e.g., the meaning of “electron spin” appears to be inextricably intertwined with the characteristic ideas of quantum theory. On the other hand, there are also terms whose meanings seem to be invariant in a number of different theories: e.g., the term “electric charge” is used in currently accepted theories of atomic structure in the same sense as in the earlier theories of Rutherford and Bohr. Similar comments apply to observation terms, however these may be specified. Accordingly, although both “theoretical” and “observational” terms may be “theory laden,” it does not follow that there can be no term in a theory which retains its meaning when it is transplanted into some other theory.

More generally, it is not clear how, on the replacement view of reduction, a theory *T* can be at the same time more inclusive than, and also have a meaning totally different from, the theory *T'* it allegedly replaces—especially since according to Feyerabend the replacing theory will entail “that all the concepts of the preceding theory have extension zero, or . . . it introduces rules which cannot be interpreted as attributing specific properties to objects within already existing classes, but which change the system of classes itself.”¹⁴ Admittedly, some of the laws and concepts of the “wider theory” often differ from their opposite numbers in the earlier theory. But even in this case, the contrasted items may not be “incommensurable.” Thus, the periodic table classifies chemical elements on the basis of certain patterns of similarity between the properties of the elements. The description (or theoretical explanation) of those properties has undergone important changes since the periodic table was first introduced by Mendeleev. Nevertheless, though the descriptions differ, the classification of the elements has remained fairly stable, so that fluorine, chlorine, bromine, and iodine, for example, continue to be included in the same class. The new theories used in formulating the classification certainly do not entail that the concepts of the preceding ones have zero extension. But it would be difficult to understand why this is so if, because of differences between the descriptions, the descriptions were totally disparate.

Consider, for example, the argument that thermodynamics is not reducible to statistical mechanics, on the ground that (among other reasons) entropy is a statistical notion in the latter theory but not in the former one: since the meaning of the word “entropy” differs in the two theories, entropy laws in statistical mechanics are not derivable from entropy laws in thermodynamics (and in fact are said to be incompatible). Admittedly, the connotation of the word “entropy” in each of the two theories is not

identical; and if the correspondence account of reduction were to claim that they are the same, it would be patently mistaken. But the fact remains that the two theories deal with many phenomena common to both their ranges; and the question is how is this possible? In brief, the answer seems to be as follows. The word "entropy" in thermodynamics is so defined that its legitimate application is limited to physical systems satisfying certain specified conditions, e.g., to systems such as gases, whose internal motions are not too "tumultuous" (the word is Planck's), a condition which is not satisfied in the case of Brownian motions. These conditions are relaxed in the definition of "entropy" in statistical mechanics, so that the extension of the Boltzmann notion of entropy includes the extension of the Clausius notion. In consequence, despite differences in the connotations of the two definitions, the theories within which they are formulated have a domain of application in common, even though the class of systems for which thermodynamical laws are approximately valid is more restricted than is the class for the laws of statistical mechanics. But it is surely not the case that the latter theory implies that the Clausius definition of entropy has a zero extension or that the laws of thermodynamics are valid for no physical systems whatsoever.

This difficulty of the replacement view in explaining how the "wider" theory, which allegedly replaces a "narrower" one, may nevertheless have a domain of common application, does not arise in the correspondence account of reduction. For the bridge laws upon which the latter sets great store are empirical hypotheses, not logically true statements in virtue of the connotations of the terms contained in them. Bridge laws state what relations presumably obtain between the *extensions* of their terms, so that in favorable cases laws of the "narrower" theory (with suitable qualifications about their approximate character) can be deduced from the "wider" theory, and thereby make intelligible why the two theories may have a common field of application. Accordingly, although I will not pretend that the correspondence account of reduction is free from difficulties or that I have resolved all of them, on the whole it is a more adequate analysis than any available alternative to it. . . .

■ | Notes

1. Paul Feyerabend, "Problems of Empiricism," in R. G. Colodny, ed., *Beyond the Edge of Certainty* (Englewood Cliffs: Prentice-Hall, Inc., 1965), p. 180.
2. Paul Feyerabend, "Reply to Criticism," *Boston Studies in the Philosophy of Science* 2 (1962), p. 252.
3. Paul Feyerabend. "Explanation, Reduction, and Empiricism," *Minnesota Studies in the Philosophy of Science* 3 (1962), p. 76.
4. Feyerabend. "Problems of Empiricism," p. 175.

5. Ibid., p. 213.
6. Ibid., p. 216.
7. Ibid.
8. Ibid., p. 175.
9. Ibid., p. 227.
10. *Boston Studies in the Philosophy of Science* p. 231; cf. also Feyerabend, "On the 'Meaning' of Scientific Terms," *Journal of Philosophy* 62 (1965), p. 271.
11. Feyerabend, "Explanation, Reduction and Empiricism," pp. 28–9, 59.
12. Many of them are noted by Dudley Shapere in his "Meaning and Scientific Change," in R. G. Colodny, ed., *Mind and Cosmos* (Pittsburgh: University of Pittsburgh Press, 1966).
13. Feyerabend, "Problems of Empiricism," p. 21; and Feyerabend, "Explanation, Reduction, and Empiricism," p. 24.
14. Feyerabend. "On the 'Meaning' of Scientific Terms," *Journal of Philosophy* 62 (1965), p. 268.

How to Be a Good Empiricist— A Plea for Tolerance in Matters Epistemological

"Facts?" he repeated. "Take a drop more grog, Mr. Franklin, and you'll get over the weakness of believing in facts! Foul play, Sir!"

—Wilkie Collins, *The Moonstone*

1 | Contemporary Empiricism Liable to Lead to Establishment of a Dogmatic Metaphysics

Today empiricism is the professed philosophy of a good many intellectual enterprises. It is the core of the sciences, or so at least we are taught, for it is responsible both for the existence and for the growth of scientific knowledge. It has been adopted by influential schools in aesthetics, ethics, and theology. And within philosophy proper the empirical point of view has been elaborated in great detail and with even greater precision. This predilection for empiricism is due to the assumption that only a thoroughly observational procedure can exclude fanciful speculation and empty metaphysics as well as to the hope that an empiristic attitude is most liable to prevent stagnation and to further the progress of knowledge. It is the purpose of the present paper to show that empiricism in the form in which it is practiced today cannot fulfill this hope.

Putting it very briefly, it seems to me that the contemporary doctrine of empiricism has encountered difficulties, and has created contradictions which are very similar to the difficulties and contradictions inherent in some versions of the doctrine of democracy. The latter are a well-known phenomenon. That is, it is well known that essentially totalitarian measures are often advertised as being a necessary consequence of democratic principles. Even worse—it not so rarely happens that the totalitarian character of the defended measures is not explicitly stated but covered up by calling them 'democratic,' the word 'democratic' now being used in a new,

and somewhat misleading, manner. This method of (conscious or unconscious) verbal camouflage works so well that it has deceived some of the staunchest supporters of true democracy. What is not so well known is that modern empiricism is in precisely the same predicament. That is, some of the methods of modern empiricism which are introduced in the spirit of anti-dogmatism and progress are bound to lead to the establishment of a dogmatic metaphysics and to the construction of defense mechanisms which make this metaphysics safe from refutation by experimental inquiry. It is true that in the process of establishing such a metaphysics the words 'empirical' or 'experience' will frequently occur; but their sense will be as distorted as was the sense of 'democratic' when used by some concealed defenders of a new tyranny.¹ This, then, is my charge: Far from eliminating dogma and metaphysics and thereby encouraging progress, modern empiricism has found a new way of making dogma and metaphysics respectable, viz., the way of calling them 'well-confirmed theories,' and of developing a method of confirmation in which experimental inquiry plays a large though well controlled role. In this respect, modern empiricism is very different indeed from the empiricism of Galileo, Faraday, and Einstein, though it will of course try to represent these scientists as following its own paradigm of research, thereby further confusing the issue.²

From what has been said above it follows that the fight for tolerance in scientific matters and the fight for scientific progress must still be carried on. What has changed is the denomination of the enemies. They were priests, or 'school-philosophers,' a few decades ago. Today they call themselves 'philosophers of science,' or 'logical empiricists.'³ There are also a good many scientists who work in the same direction. I maintain that all these groups work against scientific progress. But whereas the former did so openly and could be easily discerned, the latter proceed under the flag of progressivism and empiricism and thereby deceive a good many of their followers. Hence, although their presence is noticeable enough they may almost be compared to a fifth column, the aim of which must be exposed in order that its detrimental effect be fully appreciated. It is the purpose of this paper to contribute to such an exposure.

I shall also try to give a positive methodology for the empirical sciences which no longer encourages dogmatic petrification in the name of experience. Put in a nutshell, the answer which this method gives to the question in the title is: You can be a good empiricist only if you are prepared to work with many alternative theories rather than with a single point of view and 'experience.' This plurality of theories must not be regarded as a preliminary stage of knowledge which will at some time in the future be replaced by the One True Theory. Theoretical pluralism is assumed to be an *essential feature* of all knowledge that claims to be objective. Nor can one rest content with a plurality which is merely abstract and which is created by denying now this and now that component

of the dominant point of view. Alternatives must rather be developed in such detail that problems already 'solved' by the accepted theory can again be treated in a new and perhaps also more detailed manner. Such development will of course take time, and it will not be possible, for example, at once to construct alternatives to the present quantum theory which are comparable to its richness and sophistication. Still, it would be very unwise to bring the process to a standstill in the very beginning by the remark that some suggested new ideas are undeveloped, general, metaphysical. *It takes time to build a good theory* (a triviality that seems to have been forgotten by some defenders of the Copenhagen point of view of the quantum theory);* and it also takes time to develop an alternative to a good theory. The *function* of such concrete alternatives is, however, this: They provide means of criticizing the accepted theory in a manner which goes *beyond* the criticism provided by a comparison of that theory 'with the facts': however closely a theory seems to reflect the facts, however universal its use, and however necessary its existence seems to be to those speaking the corresponding idiom, its factual adequacy can be asserted only *after* it has been confronted with alternatives *whose invention and detailed development must therefore precede any final assertion of practical success and factual adequacy*. This, then, is the methodological justification of a plurality of *theories*: Such a plurality allows for a much sharper criticism of accepted ideas than does the comparison with a domain of 'facts' which are supposed to sit there independently of theoretical considerations. The function of unusual *metaphysical* ideas which are built up in a nondogmatic fashion and which are then developed in sufficient detail to give an (alternative) account even of the most common experimental and observational situations is defined accordingly: They play a decisive role in the

* Feyerabend is referring to the instrumentalist interpretation of quantum mechanics advocated by Niels Bohr, Werner Heisenberg, Max Born, and other pioneers of quantum physics in the 1920s. Named after the city in which Bohr worked, the Copenhagen interpretation remains the orthodox textbook account of quantum mechanics to this day. Its elements include Bohr's doctrine of complementarity, the insistence that the wave function contains a complete description of reality, and a profound reluctance to say anything about the properties of quantum systems when they are not being observed or measured. Feyerabend was a lifelong critic of "the Copenhagen mafia" (as he sometimes referred to them). As Feyerabend explains later in this reading, he rejected as unwarranted dogma Heisenberg's insistence that, while limited by the uncertainty principle, classical physics must always be used to describe the result of any experiment. See Werner Heisenberg, *Physics and Philosophy* (New York: Harper and Row, 1958); David Bohm, *Causality and Chance in Modern Physics* (New York: Harper and Row, 1961); Dugald Murdoch, *Niels Bohr's Philosophy of Physics* (Cambridge: Cambridge University Press, 1987); and James T. Cushing, *Quantum Mechanics: Historical Contingency and the Copenhagen Hegemony* (Chicago: University of Chicago Press, 1994) for good accounts of the Copenhagen interpretation and alternatives to it.

criticism and in the development of what is generally believed and 'highly confirmed'; and they have therefore to be present at *any* stage of the development of our knowledge.⁴ A science that is free from *metaphysics* is on the best way to become a *dogmatic* metaphysical system. So far the summary of the method I shall explain, and defend, in the present paper.

It is clear that this method still retains an essential element of *empiricism*: The decision between alternative theories is based upon *crucial experiments*. At the same time it must *restrict* the range of such experiments. Crucial experiments work well with theories of a low degree of generality whose principles do not touch the principles on which the ontology of the chosen observation language is based. They work well if such theories are compared with respect to a much more general background theory which provides a stable meaning for the observation sentences. However, this background theory, like any other theory, is itself in need of criticism. Criticism must use alternatives. Alternatives will be the more efficient the more radically they differ from the point of view to be investigated. It is bound to happen, then, that the alternatives do not share a single statement with the theories they criticize. Clearly, a crucial experiment is now impossible. It is impossible, not because the experimental device is too complex, or because the calculations leading to the experimental prediction are too difficult; it is impossible because there is no statement capable of expressing what emerges from the observation. This consequence, which severely restricts the domain of empirical discussion, cannot be circumvented by any of the methods which are currently in use and which all try to work with relatively stable observation languages. It indicates that the attempt to make empiricism a universal basis of all our factual knowledge cannot be carried out. The discussion of this situation is beyond the scope of the present paper.

On the whole, the paper is a concise summary of results which I have explained in a more detailed fashion in the following essays: "Explanation, Reduction, and Empiricism"; "Problems of Microphysics"; "Problems of Empiricism"; "Linguistic Philosophy and the Mind-Body Problem."⁵ All the relevant acknowledgements can be found there. Let me only repeat here that my general outlook derives from the work of K. R. Popper (London) and David Bohm (London) and from my discussions with both. It was severely tested in discussion with my colleague, T. S. Kuhn (Berkeley). It was the latter's skillful defense of a scientific conservatism which triggered two papers, including the present one. Criticism by A. Naess (Oslo), D. Rynin (Berkeley), Roy Edgley (Bristol), and J. W. N. Watkins (London) have been responsible for certain changes I made in the final version.

2 | Two Conditions of Contemporary Empiricism

In this section I intend to give an outline of some assumptions of contemporary empiricism which have been widely accepted. It will be shown in the sections to follow that these apparently harmless assumptions which have been explicitly formulated by some logical empiricists, but which also seem to guide the work of a good many physicists, are bound to lead to exactly the results I have outlined above: dogmatic petrification and the establishment, on so-called 'empirical grounds,' of a rigid metaphysics.

One of the cornerstones of contemporary empiricism is its *theory of explanation*. This theory is an elaboration of some simple and very plausible ideas first proposed by Popper⁶ and it may be introduced as follows: Let T and T' be two different scientific theories, T' the theory to be explained, or the explanandum, T the explaining theory, or the explanans. Explanation (of T') consists in the *derivation* of T' from T and initial conditions which specify the domain D' in which T' is applicable.* *Prima facie*, this demand of derivability seems to be a very natural one to make for "otherwise the explanans would not constitute adequate grounds for the explanation".⁷ It implies two things: first, that the consequences of a satisfactory explanans, T , inside D' must be compatible with the explanandum, T' ; and secondly, that the main descriptive terms of these consequences must either coincide, with respect to their meanings, with the main descriptive terms of T' , or at least they must be related to them via an empirical hypothesis. The latter result can also be formulated by saying that the meaning of T' must be unaffected by the explanation. "It is of the utmost importance," writes Professor Nagel,⁸ emphasizing this point, "that the expressions peculiar to a science will possess meanings that are fixed by its own procedures, and are therefore intelligible in terms of its own rules of usage, whether or not the science has been, or will be [explained in terms of] the other discipline."

Now if we take it for granted that more general theories are always introduced with the purpose of explaining the existent successful theories, then every new theory will have to satisfy the two conditions just mentioned. Or, to state it in a more explicit manner,

- 1 only such theories are then admissible in a given domain which either *contain* the theories already used in this domain, or which are at least *consistent* with them inside the domain⁹; and
- 2 meanings will have to be invariant with respect to scientific progress; that is, all future theories will have to be phrased in such a

* Throughout this paper, Feyerabend uses T to stand for the reducing (explaining) theory, and T' for the theory that is reduced (explained). This is exactly the opposite of Nagel's notation: for Nagel it is T' that reduces (explains) T .

manner that their use in explanations does not affect what is said by the theories, or factual reports to be explained.

These two conditions I shall call the *consistency condition* and the *condition of meaning invariance*, respectively.

Both conditions are *restrictive* conditions and therefore bound profoundly to influence the growth of knowledge. I shall soon show that the development of actual science very often violates them and that it violates them in exactly those places where one would be inclined to perceive a tremendous progress of knowledge. I shall also show that neither condition can be justified from the point of view of a tolerant empiricism. However, before doing so I would like to mention that both conditions have occasionally entered the domain of the sciences and have been used here in attacks against new developments and even in the process of theory construction itself. Especially today, they play a very important role in the construction as well as in the defense of certain points of view in microphysics.

Taking first an earlier example, we find that in his *Wärmelehre*, Ernst Mach¹⁰ makes the following remark:

Considering that there is, in a purely mechanical system of absolutely elastic atoms no real analogue for the *increase of entropy*, one can hardly suppress the idea that a violation of the second law . . . should be possible if such a mechanical system were the *real* basis of thermodynamic processes.

And referring to the fact that the second law is a highly confirmed physical law, he insinuates (in his *Zwei Aufsätze*¹¹) that for this reason the mechanical hypothesis must not be taken too seriously. There were many similar objections against the kinetic theory of heat.¹² More recently, Max Born has based his arguments against the possibility of a return to determinism upon the consistency condition and the assumption which we shall here take for granted, that wave mechanics is incompatible with determinism.

If any future theory should be deterministic it cannot be a modification of the present one, but must be entirely different. How this should be possible without sacrificing a whole treasure of well established results [i.e., without contradicting highly confirmed physical laws and thereby violating the consistency condition] I leave the determinist to worry about.¹³

Most members of the so-called Copenhagen school of quantum theory would argue in a similar manner. For them the idea of complementarity and the formalism of quantization expressing this idea do not contain any hypothetical element as they are "uniquely determined by the facts."¹⁴ Any theory which contradicts this idea is factually inadequate and must

be removed. Conversely, an explanation of the idea of complementarity is acceptable only if it either contains this idea, or is at least consistent with it. This is how the consistency condition is used in arguments against theories such as those of Bohm, de Broglie, and Vigier.¹⁵

The use of the consistency condition is not restricted to such general remarks, however. A decisive part of the existing quantum theory *itself*, viz., the projection postulate,¹⁶ is the result of the attempt to give an account of the definiteness of macro objects and macro events that is in accordance with the consistency condition. The influence of the condition of meaning invariance goes even further.

The Copenhagen interpretation of the quantum theory [writes Heisenberg¹⁷] starts from a paradox. Any experiment in physics, whether it refers to the phenomena of daily life or to atomic events, is to be described in the terms of classical physics. . . . *We cannot and should not replace these concepts by any others* [my italics]. Still the application of these concepts is limited by the relations of uncertainty. We must keep in mind this limited range of applicability of the classical concepts while using them, but we cannot and should not try to improve them.

This means that the meaning of the classical terms must remain invariant with respect to any future explanation of microphenomena. Microtheories have to be formulated in such a manner that this invariance is guaranteed. The principle of correspondence and the formalism of quantization connected with it were explicitly devised for satisfying this demand. Altogether, the quantum theory seems to be the first theory after the downfall of the Aristotelian physics that has been quite explicitly constructed with an eye both on the consistency condition and the condition of (empirical) meaning invariance. In this respect it is very different indeed from, say, relativity which violates both consistency and meaning invariance with respect to earlier theories. Most of the arguments used for the defense of its customary interpretation also depend on the validity of these two conditions and they will collapse with their removal. An examination of these conditions is therefore very topical and bound deeply to affect present controversies in microphysics. I shall start this investigation by showing that some of the most interesting developments of physical theory in the past have violated both conditions.

3 | These Conditions Not Invariably Accepted by Actual Science

The case of the consistency condition can be dealt with in a few words: it is well known (and has also been shown in great detail by Duhem¹⁸)

that Newton's theory is inconsistent with Galileo's law of the free fall and with Kepler's laws;* that statistical thermodynamics is inconsistent with the second law of the phenomenological theory; that wave optics is inconsistent with geometrical optics; and so on. Note that what is being asserted here is *logical* inconsistency; it may well be that the differences of prediction are too small to be detectable by experiment. Note also that what is being asserted is not the inconsistency of, say, Newton's theory and Galileo's law, but rather the inconsistency of *some consequences* of Newton's theory in the domain of validity of Galileo's law, and Galileo's law. In this last case the situation is especially clear. Galileo's law asserts that the acceleration of the free fall is a constant, whereas application of Newton's theory to the surface of the earth gives an acceleration that is not a constant but *decreases* (although imperceptibly) with the distance from the center of the earth. Conclusion: If actual scientific procedure is to be the measure of method, then the consistency condition is inadequate.

The case of meaning invariance requires a little more argument, not because it is intrinsically more difficult, but because it seems to be much more closely connected with deep-rooted prejudices. Assume that an explanation is required, in terms of the special theory of relativity, of the classical conservation of mass in all reactions in a closed system S . If m' , m'' , m''' , . . . , m^i , . . . are the masses of the parts P' , P'' , P''' , . . . , P^i , . . . of S , then what we want is an explanation of

$$\Sigma m' = \text{const.} \quad (1)$$

for all reactions inside S . We see at once that the consistency condition cannot be fulfilled: According to special relativity $\Sigma m'$ will vary with the velocities of the parts relative to the coordinate system in which the observations are carried out, and the total mass of S will also depend on the relative potential energies of the parts. However, if the velocities and the mutual forces are not too large, then the variation of $\Sigma m'$ predicted by relativity will be so small as to be undetectable by experiment. Now let us turn to the *meanings* of the terms in the relativistic law and in the corresponding classical law. The first indication of a possible change of meaning may be seen in the fact that in the classical case the mass of an aggregate of parts equals the sum of the masses of the parts:

$$M(\Sigma P^i) = \Sigma M(P^i).$$

* See section 4 of the reading "Physical Theory and Experiment" in chapter 3 and the section "Duhem's Critique of Inductivism: The Attack on Newtonian Method" in the accompanying commentary.

This is not valid in the case of relativity where the relative velocities and the relative potential energies contribute to the mass balance. That the relativistic concept and the classical concept of mass are very different indeed becomes clear if we also consider that the former is a *relation*, involving relative velocities, between an object and a coordinate system, whereas the latter is a *property* of the object itself and independent of its behavior in coordinate systems. True, there have been attempts to give a relational analysis even of the classical concept (Mach). None of these attempts, however, leads to the relativistic idea with its velocity dependence on the coordinate system, which idea must therefore be added even to a *relational* account of classical mass. The attempt to identify the classical mass with the relativistic rest mass is of no avail either. For although both may have the same numerical value, the one is still dependent on the coordinate system chosen (in which it is at rest and has that specific value), whereas the other is not so dependent. We have to conclude, then, that $(m)_c$ and $(m)_r$ mean very different things and that $(\Sigma m')_c = \text{const.}$ and $(\Sigma m')_r = \text{const.}$ are very different assertions. This being the case, the derivation from relativity of either equation (1) or of a law that makes slightly different quantitative predictions with $\Sigma m'$ used in the classical manner, will be possible only if a further premise is added which establishes a relation between the $(m)_c$ and the $(m)_r$. Such a 'bridge law'—and this is a major point in Nagel's theory of reduction—is a hypothesis

according to which the occurrence of the properties designated by some expression in the premises of the [explanans] is a sufficient, or a necessary and sufficient condition for the occurrence of the properties designated by the expressions of the [explanandum].¹⁹

Applied to the present case this would mean the following: Under certain conditions the occurrence of relativistic mass of a given magnitude is accompanied by the occurrence of classical mass of a corresponding magnitude; this assertion is inconsistent with another part of the explanans, viz., the theory of relativity. After all, this theory asserts that there are no invariants which are directly connected with mass measurements and it thereby asserts that ' $(m)_c$ ' does not express real features of physical systems. Thus we inevitably arrive at the conclusion that mass conservation cannot be explained in terms of relativity (or 'reduced' to relativity) without a violation of meaning invariance. And if one retorts, as has been done by some critics of the ideas expressed in the present paper,²⁰ that meaning invariance is an essential part of both reduction and explanation, then the answer will simply be that equation (1) can neither be explained by, nor reduced to relativity. Whatever the words used for describing the situation, the *fact* remains that actual science does not observe the requirement of meaning invariance.

This argument is quite general and is independent of whether the

terms whose meaning is under investigation are observable or not. It is therefore stronger than may seem at first sight. There are some empiricists who would admit that the meaning of theoretical terms may be changed in the course of scientific progress. However, not many people are prepared to extend meaning *variance* to observational terms also. The idea motivating this attitude is, roughly, that the meaning of observational terms is uniquely determined by the procedures of observation such as looking, listening, and the like. These procedures remain unaffected by theoretical advance.²¹ Hence, observational meanings, too, remain unaffected by theoretical advance. What is overlooked, here, is that the 'logic' of the observational terms is not exhausted by the procedures which are connected with their application 'on the basis of observation.' As will turn out later, it also depends on the more general ideas that determine the 'ontology' (in Quine's sense) of our discourse. These general ideas may change without any change of observational procedures being implied. For example, we may change our ideas about the nature, or the ontological status (property, relation, object, process, etc.) of the color of a self-luminescent object without changing the methods of ascertaining that color (looking, for example). Clearly, such a change is bound profoundly to influence the meanings of our observational terms.

All this has a decisive bearing upon some contemporary ideas concerning the interpretation of scientific theories. According to these ideas, theoretical terms receive their meanings via correspondence rules which connect them with an observational language *that has been fixed in advance* and independently of the structure of the theory to be interpreted. Now, our above analysis would seem to show that *if we interpret scientific theories in the manner accepted by the scientific community*, then most of these correspondence rules will be either false, or nonsensical. They will be *false* if they *assert* the existence of entities denied by the theory; they will be *nonsensical* if they *presuppose* this existence. Turning the argument around, we can also say that the attempt to interpret the calculus of some theory that has been voided of the meaning assigned to it by the scientific community with the help of the double language system, will lead to a very different theory. Let us again take the theory of relativity as an example: It can be safely assumed that the physical thing language of Carnap, and any similar language that has been suggested as an observation language, is not Lorentz-invariant. The attempt to interpret the *calculus* of relativity on *its* basis therefore cannot lead to the *theory* of relativity as it was understood by Einstein. What we shall obtain will be at the very most *Lorentz's interpretation* with its inherent asymmetries. This undesirable result cannot be evaded by the *demand* to use a different and more adequate observation language. The double language system assumes that theories which are not connected with some observation language do not possess an interpretation. The demand assumes that they do, and asks to choose the observation language most suited to it. It reverses the relation

between theory and experience that is characteristic for the double language method of interpretation, which means, it gives up this method. Contemporary empiricism, therefore, has not led to any satisfactory account of the meanings of scientific theories.²²

What we have shown so far is that the two conditions of Section 2 are frequently violated in the course of scientific practice and especially at periods of scientific revolution. This is not yet a very strong argument. True: There are empirically inclined philosophers who have derived some satisfaction from the assumption that they only make explicit what is implicitly contained in scientific practice. It is therefore quite important to show that scientific practice is not what it is supposed to be by them. Also, strict adherence to meaning invariance and consistency would have made impossible some very decisive advances in physical theory such as the advance from the physics of Aristotle to the physics of Galileo and Newton. However, how do we know (independently of the fact that they do exist, have a certain structure, and are very influential—a circumstance that will have great weight with opportunists only²³) that the sciences are a desirable phenomenon, that they contribute to the advancement of knowledge, and that their analysis will therefore lead to reasonable methodological demands? And did it not emerge in the last section that meaning invariance and the consistency condition *are* adopted by some scientists? Actual scientific practice, therefore, cannot be our last authority. We have to find out whether consistency and meaning invariance are *desirable* conditions and this quite independently of who accepts and praises them and how many Nobel prizes have been won with their help.²⁴ Such an investigation will be carried out in the next sections.

4 | Inherent Unreasonableness of Consistency Condition

Prima facie, the case of the consistency condition can be dealt with in very few words. Consider for that purpose a theory T' that successfully describes the situation in the domain D' . From this we can infer (a) that T' agrees with a *finite* number of observations (let their class be F); and (b) that it agrees with these observations inside a margin M of error only.²⁵ Any alternative that contradicts T' outside F and inside M is supported by exactly the same observations and therefore acceptable if T' was acceptable (we shall assume that F are the only observations available). The consistency condition is much less tolerant. It eliminates a theory not because it is in disagreement with the *facts*; it eliminates it because it is in disagreement with *another theory*, with a theory, moreover, whose confirming instances it shares. *It thereby makes the as yet untested part of that theory a measure of validity.* The only difference between such a measure and a more recent theory is age and familiarity. Had the younger theory been

there first, then the consistency condition would have worked in its favor. In this respect the effect of the consistency condition is rather similar to the effect of the more traditional methods of transcendental deduction, analysis of essences, phenomenological analysis, linguistic analysis. It contributes to the preservation of the old and familiar not because of any inherent advantage in it—for example, not because it has a better foundation in observation than has the newly suggested alternative, or because it is more elegant—but just because it is old and familiar. This is not the only instance where on closer inspection a rather surprising similarity emerges between modern empiricism and some of the school philosophies it attacks.

Now it seems to me that these brief considerations, although leading to an interesting *tactical* criticism of the consistency condition, do not yet go to the heart of the matter. They show that an alternative of the accepted point of view which shares its confirming instances cannot be *eliminated* by factual reasoning. They do not show that such an alternative is *acceptable*; and even less do they show that it *should be used*. It is bad enough, so a defender of the consistency condition might point out, that the accepted point of view does not possess full empirical support. Adding new theories of an *equally unsatisfactory character* will not improve the situation; nor is there much sense in trying to *replace* the accepted theories by some of their possible alternatives. Such replacement will be no easy matter. A new formalism may have to be learned and familiar problems may have to be calculated in a new way. Textbooks must be rewritten, university curricula readjusted, experimental results reinterpreted. And what will be the result of all the effort? Another theory which, from an empirical point of view, has no advantage whatever over and above the theory it replaces. The only real improvement, so the defender of the consistency condition will continue, derives from the *addition of new facts*. Such new facts will either support the current theories, or they will force us to modify them by indicating precisely where they go wrong. In both cases they will precipitate real progress and not only arbitrary change. The proper procedure must therefore consist in the confrontation of the accepted point of view with as many relevant facts as possible. The exclusion of alternatives is then required for reasons of expediency: Their invention not only does not help, but it even hinders progress by absorbing time and manpower that could be devoted to better things. And the function of the consistency condition lies precisely in this. It eliminates such fruitless discussion and it forces the scientist to concentrate on the facts which, after all, are the only acceptable judges of a theory. This is how the practicing scientist will defend his concentration on a single theory to the exclusion of all empirically possible alternatives.²⁶

It is worthwhile repeating the reasonable core of this argument: Theories should not be changed unless there are pressing reasons for doing so. The only pressing reason for changing a theory is disagreement with

facts. Discussion of incompatible facts will therefore lead to progress. Discussion of incompatible alternatives will not. Hence, it is sound procedure to increase the number of relevant facts. It is not sound procedure to increase the number of factually adequate, but incompatible alternatives. One might wish to add that formal improvements such as increase of elegance, simplicity, generality, and coherence should not be excluded. But once these improvements have been carried out, the collection of facts for the purpose of test seems indeed to be the only thing left to the scientist.

5 | Relative Autonomy of Facts

And this it is—provided these facts *exist, and are available independently of whether or not one considers alternatives to the theory to be tested*. This assumption on which the validity of the argument in the last section depends in a most decisive manner I shall call the assumption of the relative autonomy of facts, or the autonomy principle. It is not asserted by this principle that the discovery and description of facts is independent of *all* theorizing. But it is asserted that the facts which belong to the empirical content of some theory are available whether or not one considers alternatives to *this* theory. I am not aware that this very important assumption has ever been explicitly formulated as a separate postulate of the empirical method. However, it is clearly implied in almost all investigations which deal with questions of confirmation and test. All these investigations use a model in which a *single* theory is compared with a class of facts (or observation statements) which are assumed to be 'given' somehow. I submit that this is much too simple a picture of the actual situation. Facts and theories are much more intimately connected than is admitted by the autonomy principle. Not only is the description of every single fact dependent on *some* theory (which may, of course, be very different from the theory to be tested). There exist also facts which cannot be unearthed except with the help of alternatives to the theory to be tested, and which become unavailable as soon as such alternatives are excluded. This suggests that the methodological unit to which we must refer when discussing questions of test and empirical content is constituted by a *whole set of partly overlapping, factually adequate, but mutually inconsistent theories*. In the present paper only the barest outlines will be given of such a test model. However, before doing this I want to discuss an example which shows very clearly the function of alternatives in the discovery of facts.

As is well known, the Brownian particle is a perpetual motion machine of the second kind and its existence refutes the phenomenological

second law.* It therefore belongs to the domain of relevant facts for this law. Now, could this relation between the law and the Brownian particle have been discovered in a *direct* manner, i.e., could it have been discovered by an investigation of the observational consequences of the phenomenological theory that did not make use of an alternative account of heat? This question is readily divided into two: (1) Could the *relevance* of the Brownian particle have been discovered in this manner? (2) Could it have been demonstrated that it actually *refutes* the second law? The answer to the first question is that we do not know. It is impossible to say what would have happened had the kinetic theory not been considered by some physicists. It is my guess, however, that in this case the Brownian particle would have been regarded as an oddity much in the same way in which some of the late Professor Ehrenhaft's astounding effects²⁷ are regarded as an oddity, and that it would not have been given the decisive position it assumes in contemporary theory. The answer to the second question is simply—No. Consider what the discovery of the inconsistency between the Brownian particle and the second law would have required! It would have required (a) measurement of the exact *motion* of the particle in order to ascertain the changes of its kinetic energy plus the energy spent on overcoming the resistance of the fluid; and (b) it would have required precise measurements of temperature and heat transfer in the surrounding medium in order to ascertain that any loss occurring here was indeed compensated by the increase of the energy of the moving particle and the work done against the fluid. Such measurements are beyond experimental possibilities.²⁸ Neither is it possible to make precise measurements of the heat transfer; nor can the path of the particle be investigated with the desired precision. Hence a 'direct' refutation of the second law that con-

* A perpetual motion machine of the second kind—or, as it is often called, perpetual motion of the second kind—is any machine, device, or system that can produce work merely by taking heat from a body (that is, by cooling the body, transforming the heat energy into work, and producing no other effect). The pioneers of classical thermodynamics (Clausius, Kelvin) took the impossibility of such a machine to be the empirical foundation of the second law of thermodynamics, just as the impossibility of perpetual motion of the first kind—the production of work without the transformation of energy—was regarded as the foundation of the first law (energy conservation). The second law asserts that entropy can never decrease in a isolated system. Brownian particles violate the second law because their random motion is acquired at the expense of very brief spontaneous decreases of entropy in the fluid in which they are suspended. (Even if the definition of perpetual motion of the second kind were to stipulate that the machine must operate continuously, in a cycle, it is still possible that Brownian particles might satisfy that definition, though the probability that their random motions acquire the necessary degree of coordination, purely by chance, is extraordinarily low. Because this probability is so very low, Brownian motion cannot be harnessed to produce usable work.)

siders only the phenomenological theory and the 'facts' of Brownian motion is impossible. And, as is well known, the actual refutation was brought about in a very different manner. It was brought about via the kinetic theory and Einstein's utilization of it in the calculation of the statistical properties of the Brownian motion.²⁹ In the course of this procedure the phenomenological theory (T') was incorporated into the wider context of statistical physics (T) *in such a manner that the consistency condition was violated*; and *then* a crucial experiment was staged (investigations of Svedberg and Perrin).

It seems to me that this example is typical for the relation between fairly general theories, or points of view, and 'the facts.' Both the relevance and the refuting character of many very decisive facts can be established only with the help of other theories which, although factually adequate, are yet not in agreement with the view to be tested. This being the case, the production of such refuting facts may have to be preceded by the invention and articulation of alternatives to that view. Empiricism demands that the empirical content of whatever knowledge we possess be increased as much as possible. Hence *the invention of alternatives in addition to the view that stands in the center of discussion constitutes an essential part of the empirical method*. Conversely, the fact that the consistency condition eliminates alternatives now shows it to be in disagreement with empiricism and not only with scientific practice. By excluding valuable tests it decreases the empirical content of the theories which are permitted to remain (and which, as we have indicated above, will usually be the theories which have been there first); and it especially decreases the number of those facts which could show their limitations. This last result of a determined application of the consistency condition is of very topical interest. It may well be that the refutation of the quantum-mechanical uncertainties presupposes just such an incorporation of the present theory into a wider context which is no longer in accordance with the idea of complementarity and which therefore suggests new and decisive experiments. And it may also be that the insistence, on the part of the majority of contemporary physicists, on the consistency condition will, if successful, forever protect these uncertainties from refutation. This is how modern empiricism may finally lead to a situation where a certain point of view petrifies into dogma by being, in the name of experience, completely removed from any conceivable criticism.

6 | The Self-Deception Involved in all Uniformity

It is worthwhile to examine this apparently empirical defense of a dogmatic point of view in somewhat greater detail. Assume that physicists have adopted, either consciously or unconsciously, the idea of the uniqueness

of complementarity and that they therefore elaborate the orthodox point of view and refuse to consider alternatives. In the beginning such a procedure may be quite harmless. After all, a man can do only so many things at a time and it is better when he pursues a theory in which he is interested rather than a theory he finds boring. Now assume that the pursuit of the theory he chose has led to successes and that the theory has explained in a satisfactory manner circumstances that had been unintelligible for quite some time. This gives empirical support to an idea which to start with seemed to possess only this advantage: It was interesting and intriguing. The concentration upon the theory will now be reinforced, the attitude towards alternatives will become less tolerant. Now if it is true, as has been argued in the last section, that many facts become available only with the help of such alternatives, then the refusal to consider them *will result in the elimination of potentially refuting facts*. More especially, it will eliminate facts whose discovery would show the complete and irreparable inadequacy of the theory.³⁰ Such facts having been made inaccessible, the theory will appear to be free from blemish and it will seem that "all evidence points with merciless definiteness in the . . . direction . . . [that] all the processes involving . . . unknown interactions conform to the fundamental quantum law".³¹ This will further reinforce the belief in the uniqueness of the current theory and in the complete futility of any account that proceeds in a different manner. Being now very firmly convinced that there is only one good microphysics, the physicists will try to explain even adverse facts in its terms, and they will not mind when such explanations are sometimes a little clumsy. By now the success of the theory has become public news. Popular science books (and this includes a good many books on the philosophy of science) will spread the basic postulates of the theory; applications will be made in distant fields. More than ever the theory will appear to possess tremendous empirical support. The chances for the consideration of alternatives are now very slight indeed. The final success of the fundamental assumptions of the quantum theory and of the idea of complementarity will seem to be assured.

At the same time it is evident, on the basis of the considerations in the last section, that this appearance of success *cannot in the least be regarded as a sign of truth and correspondence with nature*. Quite the contrary, the suspicion arises that the absence of major difficulties is a result of the decrease of empirical content brought about by the elimination of alternatives, and of facts that can be discovered with the help of these alternatives only. In other words, *the suspicion arises that this alleged success is due to the fact that in the process of application to new domains the theory has been turned into a metaphysical system*. Such a system will of course be very 'successful' not, however, because it agrees so well with the facts, but because no facts have been specified that would constitute a test and because some such facts have even been removed. Its 'success' is *entirely manmade*. It was decided to stick to some ideas and the result was,

quite naturally, the survival of these ideas. If now the initial decision is forgotten, or made only implicitly, then the survival will seem to constitute independent support, it will reinforce the decision, or turn it into an explicit one, and in this way close the circle. This is how empirical 'evidence' may be *created* by a procedure which quotes as its justification the very same evidence it has produced in the first place.

At this point an 'empirical' theory of the kind described (and let us always remember that the basic principles of the present quantum theory and especially the idea of complementarity are uncomfortably close to forming such a theory) becomes almost indistinguishable from a myth. In order to realize this, we need only consider that on account of its all-pervasive character a myth such as the myth of witchcraft and of demonic possession will possess a high degree of confirmation on the basis of observation. Such a myth has been taught for a long time; its content is enforced by fear, prejudice, and ignorance as well as by a jealous and cruel priesthood. It penetrates the most common idiom, infects all modes of thinking and many decisions which mean a great deal in human life. It provides models for the explanation of any conceivable event, conceivable, that is, for those who have accepted it.³² This being the case, its key terms will be fixed in an unambiguous manner and the idea (which may have led to such a procedure in the first place) that they are copies of unchanging entities and that change of meaning, if it should happen, is due to human mistake—this idea will now be very plausible. Such plausibility reinforces all the manoeuvres which are used for the preservation of the myth (elimination of opponents included). The conceptual apparatus of the theory and the emotions connected with its application having penetrated all means of communication, all actions, and indeed the whole life of the community, such methods as transcendental deduction, analysis of usage, phenomenological analysis which are means for further solidifying the myth will be extremely successful (which shows, by the way, that all these methods which have been the trademark of various philosophical schools old and new, have one thing in common: They tend to *preserve* the *status quo* of the intellectual life).³³ Observational results too, will speak in favor of the theory as they are formulated in its terms. It will seem that at last the truth has been arrived at. At the same time it is evident that all contact with the world has been lost and that the stability achieved, the semblance of absolute truth, is *nothing but the result of an absolute conformism*.³⁴ For how can we possibly test, or improve upon, the truth of a theory if it is built in such a manner that any conceivable event can be described, and explained, in terms of its principles? The *only* way of investigating such all-embracing principles is to compare them with a different set of *equally all-embracing* principles—but this way has been excluded from the very beginning. The myth is therefore of no objective relevance, it continues to exist solely as the result of the effort of the community of believers and of their leaders, be these now priests or Nobel

prize winners. *Its 'success' is entirely manmade.* This, I think, is the most decisive argument against any method that encourages uniformity, be it now empirical or not. Any such method is in the last resort a method of deception. It enforces an unenlightened conformism, and speaks of truth; it leads to a deterioration of intellectual capabilities, of the power of imagination, and speaks of deep insight; it destroys the most precious gift of the young, their tremendous power of imagination, and speaks of education.

To sum up: *Unanimity of opinion may be fitting for a church, for the frightened victims of some (ancient, or modern) myth, or for the weak and willing followers of some tyrant; variety of opinion is a feature necessary for objective knowledge; and a method that encourages variety is also the only method that is compatible with a humanitarian outlook.* To the extent to which the consistency condition (and, as will emerge, the condition of meaning invariance) delimits variety, it contains a theological element (which lies, of course, in the worship of 'facts' so characteristic for nearly all empiricism).

7 | Inherent Unreasonableness of Meaning Invariance

What we have achieved so far has immediate application to the question whether the meaning of certain key terms should be kept unchanged in the course of the development and improvement of our knowledge. After all, the meaning of every term we use depends upon the theoretical context in which it occurs. Hence, if we consider two contexts with basic principles which either contradict each other, or which lead to inconsistent consequences in certain domains, it is to be expected that some terms of the first context will not occur in the second context with exactly the same meaning. Moreover, if our methodology demands the use of mutually inconsistent, partly overlapping, and empirically adequate theories, then it thereby also demands the use of conceptual systems which are mutually *irreducible* (their primitives cannot be connected by bridge laws which are meaningful *and* factually correct) and it demands that meanings of terms be left elastic and that no binding commitment be made to a certain set of concepts.

It is very important to realize that such a tolerant attitude towards meanings, or such a change of meaning in cases where one of the competing conceptual systems has to be abandoned need not be the result of directly accessible observational difficulties. The law of inertia of the so-called *impetus theory* of the later Middle Ages³⁵ and Newton's own law of inertia are in perfect quantitative agreement: Both assert that an object that is not under the influence of any outer force will proceed along a straight line with constant speed. Yet despite this fact, the adoption of

Newton's theory entails a conceptual revision that forces us to abandon the inertial law of the impetus theory, not because it is quantitatively incorrect but *because it achieves the correct predictions with the help of inadequate concepts*. The law asserts that the *impetus* of an object that is beyond the reach of outer forces remains constant.³⁶ The impetus is interpreted as an inner *force* which pushes the object along. Within the impetus theory such a force is quite conceivable as it is assumed here that forces determine *velocities* rather than accelerations. The concept of impetus is therefore formed in accordance with a law (forces determine velocities) and this law is inconsistent with the laws of Newton's theory and must be abandoned as soon as the latter is adopted. This is how the progress of our knowledge may lead to conceptual revisions for which no direct observational reasons are available. The occurrence of such changes quite obviously refutes the contention of some philosophers that the invariance of *usage* in the trivial and uninteresting contexts of the private lives of not too intelligent and inquisitive people indicates invariance of *meaning* and the superficiality of all scientific changes. It is also a very decisive objection against any crudely operationalistic account of both observable terms and theoretical terms.

What we have said applies even to singular statements of observation. Statements which are empirically adequate, and which are the result of observation (such as 'here is a table') may have to be reinterpreted, not because it has been found that they do not adequately express what is seen, heard, felt, but because of some changes in sometimes very remote parts of the conceptual scheme to which they belong. Witchcraft is again a very good example. Numerous eyewitnesses claim that they have actually *seen* the devil, or *experienced* demonic influence. There is no reason to suspect that they were lying. Nor is there any reason to assume that they were sloppy observers, for the phenomena leading to the belief in demonic influence are so obvious that a mistake is hardly possible (possession; split personality; loss of personality; hearing voices; etc.). These phenomena are well known today.³⁷ In the conceptual scheme that was the one generally accepted in the 15th and 16th centuries, the only way of describing them, or at least the way that seemed to express them most adequately, was by reference to demonic influences. Large parts of this conceptual scheme were changed for philosophical reasons and also under the influence of the evidence accumulated by the sciences. Descartes' materialism played a very decisive role in discrediting the belief in spatially localizable spirits. The language of demonic influences was no part of the new conceptual scheme that was created in this manner. It was for this reason that a reformulation was needed, and a reinterpretation of even the most common 'observational' statements. Combining this example with the remarks at the beginning of the present section, we now realize that according to the method of classes of alternative theories a lenient attitude must be taken with respect to the meanings of all the terms we use. We must not

attach too great an importance to 'what we mean' by a phrase, and we must be prepared to change whatever little we have said concerning this meaning as soon as the need arises. Too great concern with meanings can only lead to dogmatism and sterility. Flexibility, and even sloppiness in semantical matters is a prerequisite of scientific progress.³⁸

8 | Some Consequences

Three consequences of the results so far obtained deserve a more detailed discussion. The first consequence is an evaluation of *metaphysics* which differs significantly from the standard empirical attitude. As is well known, there are empiricists who demand that science start from observable facts and proceed by generalization, and who refuse the admittance of metaphysical ideas at any point of this procedure. For them, only a system of thought that has been built up in a purely inductive fashion can claim to be genuine knowledge. Theories which are partly metaphysical, or 'hypothetical,' are suspect, and are best not used at all. This attitude has been formulated most clearly by Newton³⁹ in his reply to Pardies' second letter concerning the theory of colors:

if the possibility of hypotheses is to be the test of truth and reality of things, I see not how certainty can be obtained in any science; since numerous hypotheses may be devised, which shall seem to overcome new difficulties.

This radical position, which clearly depends on the demand for a theoretical monism, is no longer as popular as it used to be. It is now granted that metaphysical considerations may be of importance when the task is to *invent* a new physical theory; such invention, so it is admitted, is a more or less irrational act containing the most diverse components. Some of these components are, and perhaps must be, metaphysical ideas. However, it is also pointed out that as soon as the theory has been developed in a formally satisfactory fashion and has received sufficient confirmation to be regarded as empirically successful, it is pointed out that in the very same moment it can *and must* forget its metaphysical past; metaphysical speculation must *now* be replaced by empirical argument.

On the one side I would like to emphasize [writes Ernst Mach on this point⁴⁰] that *every and any* idea is admissible as a means for research, provided it is helpful; still, it must be pointed out, on the other side, that it is very necessary from time to time to free the presentation of the *results* of research from all inessential additions.

This means that empirical considerations are still given the upper hand over metaphysical reasoning. Especially in the case of an inconsistency between metaphysics and some highly confirmed empirical theory it will be decided, *as a matter of course*, that the theory or the result of observation must stay, and that the metaphysical system must go. A very simple example is the way in which materialism is being judged by some of its opponents. For a materialist the world consists of material particles moving in space, of collections of such particles. Sensations, as introspected by human beings, do not look like collections of particles, and their observed existence is therefore assumed to refute and thereby to remove the metaphysical doctrine of materialism. Another example which I have analyzed in "Problems of Microphysics" is the attempt to eliminate certain very general ideas concerning the nature of microentities on the basis of the remark that they are inconsistent "with an immense body of experience" and that "to object to a lesson of experience by appealing to metaphysical preconceptions is unscientific."⁴¹

The methodology developed in the present paper leads to a very different evaluation of metaphysics. Metaphysical systems are scientific theories in their most primitive stage. If they *contradict* a well-confirmed point of view, then this indicates their usefulness as an alternative to this point of view. Alternatives are needed for the purpose of criticism. Hence, metaphysical systems which contradict observational results or well-confirmed theories *are most welcome* starting points of such criticism. Far from being misfired attempts at anticipating, or circumventing, empirical research which were deservedly exposed by a reference to experience, they are the only means at our disposal for examining those parts of our knowledge which have already become observational and which are therefore inaccessible to a criticism 'on the basis of observation.'

A second consequence is that a new attitude has to be adopted with respect to the *problem of induction*. This problem consists in the question of what justification there is for asserting the truth of a statement *S* given the truth of another statement, *S'*, whose content is smaller than the content of *S*. It may be taken for granted that those who want to justify the truth of *S* also assume that after the justification the truth of *S* will be *known*. Knowledge to the effect that *S* implies the *stability* of *S* (we must not change, remove, criticize, what we know to be true). The method we are discussing at the present moment cannot allow such stability. It follows that the problem of induction at least in some of its formulations, is a problem whose solution leads to undesirable results. It may therefore be properly termed a *pseudo problem*.

The third consequence, which is more specific, is that *arguments from synonymy* (or from coextensionality), far from being that measure of adequacy as which they are usually introduced, are liable severely to impede the progress of knowledge. Arguments from synonymy judge a theory or a point of view not by its capability to mimic the world but rather by its

capability to mimic the descriptive terms of another point of view which for some reason is received favorably. Thus for example, the attempt to give a materialistic, or else a purely physiological, account of human beings is criticized on the grounds that materialism, or physiology, cannot provide synonyms for 'mind,' 'pain,' 'seeing red,' 'thinking of Vienna,' in the sense in which these terms are used either in ordinary English (provided there is a well-established usage concerning these terms, a matter which I doubt) or in some more esoteric mentalistic idiom. Clearly, such criticism silently assumes the principle of meaning invariance, that is, it assumes that the meanings of at least some fundamental terms must remain unchanged in the course of the progress of our knowledge. It cannot therefore be accepted as valid.⁴²

However, we can, and must go, still further. The ideas which we have developed above are strong enough not only to *reject* the demand for synonymy, wherever it is raised, but also to *support* the demand for irreducibility (in the sense in which this notion was used at the beginning of Section 7). The reason is that irreducibility is a presupposition of high critical ability on the part of the point of view shown to be irreducible. An outer indication of such irreducibility which is quite striking in the case of an attack upon commonly accepted ideas is the feeling of *absurdity*: We deem absurd what goes counter to well-established linguistic habits. The absence, from a newly introduced set of ideas, of synonymy relations connecting it with parts of the accepted point of view; the feeling of absurdity therefore indicate that the new ideas are fit for the purpose of criticism, i.e., that they are fit for either leading to a strong *confirmation* of the earlier theories, or else to a very revolutionary *discovery*: absence of synonymy, clash of meanings, absurdity are desirable. Presence of synonymy, intuitive appeal, agreement with customary modes of speech, far from being *the* philosophical virtue, indicates that not much progress has been made and that the business of investigating what is commonly accepted *has not even started*.

9 | How To Be a Good Empiricist

The final reply to the question put in the title is therefore as follows. A good empiricist will not rest content with the theory that is in the center of attention and with those tests of the theory which can be carried out in a direct manner. Knowing that the most fundamental and the most general criticism is the criticism produced with the help of alternatives, he will try to invent such alternatives.⁴³ It is, of course, impossible at once to produce a theory that is formally comparable to the main point of view and that leads to equally many predictions. His first step will therefore be the formulation of fairly general assumptions which are not yet directly

connected with observations; this means that his first step will be the invention of a new *metaphysics*. This metaphysics must then be elaborated in sufficient detail in order to be able to compete with the theory to be investigated as regards generality, details of prediction, precision of formulation.⁴⁴ We may sum up both activities by saying that a good empiricist must be a critical metaphysician. Elimination of all metaphysics, far from increasing the empirical content of the remaining theories, is liable to turn these theories into dogmas. The consideration of alternatives together with the attempt to criticize each of them in the light of experience also leads to an attitude where meanings do not play a very important role and where arguments are based upon assumptions of fact rather than analysis of (archaic, although perhaps very precise) meanings. The effect of such an attitude upon the development of human capabilities should not be underestimated either. Where speculation and invention of alternatives is encouraged, bright ideas are liable to occur in great number and such ideas may then lead to a change of even the most 'fundamental' parts of our knowledge, i.e., they may lead to a change of assumptions which either are so close to observation that their truth seems to be dictated by 'the facts,' or which are so close to common prejudice that they seem to be 'obvious,' and their negation 'absurd.' In such a situation it will be realized that neither 'facts' nor abstract ideas can ever be used for defending certain principles come what may. Wherever facts play a role in such a dogmatic defense, we shall have to suspect foul play (see the opening quotation) — the foul play of those who try to turn good science into bad, because unchangeable, metaphysics. In the last resort, therefore, being a good empiricist means being critical, and basing one's criticism not just on an abstract principle of skepticism but upon *concrete suggestions* which indicate in every single case how the accepted point of view might be further tested and further investigated and which thereby prepare the next step in the development of our knowledge.⁴⁵

■ | Notes

1. K. R. Popper, *The Open Society and Its Enemies*, Princeton University Press, Princeton, New Jersey, 1953.
2. It is very interesting to see how many so-called empiricists, when turning to the past completely fail to pay attention to some very obvious facts which are incompatible with their empiristic epistemology. Thus Galileo has been represented as a thinker who turned away from the empty speculations of the Aristotelians and who based his own laws upon facts which he had carefully collected beforehand. Nothing could be further from the truth. *The Aristotelians could quote numerous observational results in their favor.* The Copernican idea of the motion of the earth, on the other hand, did not possess independent observational support, at least not

in the first 150 years of its existence. Moreover, it was inconsistent with facts and highly confirmed physical theories. And *this* is how modern physics started: not as an observational enterprise *but as an unsupported speculation that was inconsistent with highly confirmed laws*. For details and further references see P. K. Feyerabend, "Realism and Instrumentalism—Comments on the Logic of Factual Support," in M. Bunge, Ed., *The Critical Approach to Science and Philosophy*, The Free Press, New York, 1964.

3. One might be inclined to add those who base their pronouncements upon an analysis of what they call 'ordinary language.' I do not think they deserve to be honored by a criticism. Paraphrasing Galileo, one might say that they "deserve not even that name, for they do not talk plainly and simply but are content to adore the shadows, philosophizing not with due circumspection but merely from having memorized a few ill-understood principles."

4. It is nowadays frequently assumed that "if one considers the history of a special branch of science, one gets the impression that non-scientific elements . . . relatively frequently occur in the earlier stages of development, but that they gradually retrogress in later stages and even tend to disappear in such advanced stages which become ripe for more or less thorough formalization." (H. J. Groenewold, *Synthese*, 1957, p. 305). Our considerations in the text would seem to show that such a development is very undesirable and can only result in a well-formalized, precisely expressed, and completely petrified metaphysics.

5. These essays were published in Volume III of the *Minnesota Studies in the Philosophy of Science*; in Volumes I and II of the *Pittsburgh Studies in the Philosophy of Science*; and in *Problems of Philosophy, Essays in Honor of Herbert Feigl*, respectively.

6. See K. R. Popper, *Logic of Scientific Discovery*, New York, 1959, Section 12. (This is a translation of his *Logik der Forschung* published in 1935.) The decisive feature of Popper's theory, a feature which was not at all made clear by earlier writers on the subject of explanation, is the emphasis he puts on the initial conditions and the implied possibility of two kinds of laws, viz., (1) laws concerning the temporal sequence of events; and (2) laws concerning the space of initial conditions. In the case of the quantum theory, the laws of the second kind provide very important information about the nature of the elementary particles and it is to *them* and *not* to the laws of motion that reference is made in the discussions concerning the interpretation of the uncertainty relations. In general relativity, the laws formulating the initial conditions concern the structure of the universe at large and only by overlooking them could it be believed that a purely relational account of space would be possible. For the last point, cf. E. I. Hill, "Quantum Physics and the Relativity Theory," in H. Feigl and G. Maxwell, Eds., *Current Issues in the Philosophy of Science*, Holt, Rinehart and Winston, New York, 1961.

7. C. G. Hempel, "Studies in the Logic of Explanation," reprinted in H. Feigl and M. Brodbeck, Eds., *Readings in the Philosophy of Science*, New York, 1953, p. 321.

8. E. Nagel, "The Meaning of Reduction in the Natural Sciences," reprinted in A. C. Danto and S. Morgenbesser, Eds., *Philosophy of Science*, New York, 1960, p. 301.

9. It has been objected to this formulation that theories which are consistent with a given explanandum may still contradict each other. This is quite correct, but it does not invalidate my argument. For as soon as a single theory is regarded as sufficient for explaining all that is known (and represented by the other theories in question), it will have to be consistent with all these other theories.
10. E. Mach, *Wärmelehre*, Leipzig, 1897, p. 364.
11. E. Mach, *Zwei Aufsätze*, Leipzig, 1912.
12. For a discussion of these objections, cf. ter Haar's review article in *Reviews of Modern Physics*, 1957.
13. M. Born, *Natural Philosophy of Cause and Chance*, Oxford University Press, New York, 1948, p. 109.
14. L. Rosenfeld, "Misunderstandings about the Foundations of the Quantum Theory," in S. Körner, Ed., *Observation and Interpretation*, London, 1957, p. 42.
15. Cf. the discussions in *Observation and Interpretation* (see previous note).
16. For details and further literature, cf. Section 11 of my paper "Problems of Microphysics," in R. Colodny, Ed., *Frontiers of Science and Philosophy*: University of Pittsburgh Press, Pittsburgh, 1962, pp. 189–283.
17. W. Heisenberg, *Physics and Philosophy*, New York, 1958, p. 44.
18. P. Duhem, *La Théorie Physique: Son Objet, Sa Structure*, Paris, 1914, Chapters IX and X. See also K. R. Popper, "The Aim of Science," *Ratio*, Vol. I (1957) 24–35.
19. E. Nagel, "The Meaning of Reduction in the Natural Sciences," loc. cit., p. 302.
20. Cf. Section 4.7 of M. Scriven's paper "Explanations, Predictions, and Laws," in Vol. III of the *Minnesota Studies in the Philosophy of Science*. Similar objections have been raised by Viktor Kraft (Vienna) and David Rynin (Berkeley).
21. For an exposition and criticism of this idea cf. my 'Attempt at a Realistic Interpretation of Experience,' *Proceedings of the Aristotelian Society*, New Series, LVIII, 143–170 (1958).
22. It must be admitted, however, that Einstein's original interpretation of the special theory of relativity is hardly ever used by contemporary physicists. For them the theory of relativity consists of two elements: (1) the Lorentz transformations; and (2) mass-energy equivalence. The Lorentz transformations are interpreted purely formally and are used to make a selection among possible equations. This interpretation does not allow one to distinguish between Lorentz's original point of view and the entirely different point of view of Einstein. According to it Einstein achieved a very minor *formal* advance (this is the basis of Whittaker's attempt to 'debunk' Einstein). It is also very similar to what application of the double language model would yield. Still, an undesirable philosophical procedure is not improved by the support it gets from an undesirable procedure in physics. (The above comment on the contemporary attitude towards relativity was made by E. L. Hill in discussions at the Minnesota Center for the Philosophy of Science.)
23. In about 1925 philosophers of science were bold enough to stick to their theses even in those cases where they were inconsistent with actual science. They meant

to be *reformers* of science, and not *imitators*. (This point was explicitly made by Mach in his controversy with Planck. Cf. again his *Zwei Aufsätze*.) In the meantime they have become rather tame (or beat) and are much more prepared to change their ideas in accordance with the latest discoveries of the historians, or the latest fashion of the contemporary scientific enterprise. This is very regrettable, indeed, for it considerably decreases the number of the rational critics of the scientific enterprise. And it also seems to give unwanted support to the Hegelian thesis (which is now implicitly held by many historians and philosophers of science) that what exists has a 'logic' of its own and is for that very reason reasonable.

24. Even the most dogmatic enterprise allows for discoveries (cf. the 'discovery' of so-called 'white Jews' among German physicists during the Nazi period). Hence, before hailing a so-called discovery, we must make sure that the system of thought which forms its background is not of a dogmatic kind.

25. The indefinite character of all observations has been made very clear by Duhem, *La Théorie Physique: Son Objet, Sa Structure*, Chap. IX. For an alternative way of dealing with this indefiniteness, cf. S. Körner, *Conceptual Thinking*, New York, 1960.

26. More detailed evidence for the existence of this attitude and for the way in which it influences the development of the sciences may be found in T. S. Kuhn, *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago, 1962. The attitude is extremely common in the contemporary quantum theory. 'Let us enjoy the successful theories we possess and let us not waste our time with contemplating what *would* happen if *other* theories were used'—this seems to be the motto of almost all contemporary physicists (cf. W. Heisenberg, *Physics and Philosophy*, pp. 56, 144) and philosophers (cf. N. R. Hanson, "Five Cautions for the Copenhagen Critics," *Philosophy of Science*, XXVI, 325–337 [1959]). It may be traced back to Newton's papers and letters (to Hooke, and Pardies) on the theory of color. See also [note] 23.

27. Having witnessed these effects under a great variety of conditions, I am much more reluctant to regard them as mere curiosities than is the scientific community of today. Cf. also my edition of Ehrenhaft's lectures, *Einzelne Magnetische Nord- und Südpole und deren Auswirkung in den Naturwissenschaften*, Vienna, 1947. [See Gerald Holton, "Subelectrons, Presuppositions, and the Millikan-Ehrenhaft Dispute," in *The Scientific Imagination* (Cambridge: Cambridge University Press, 1978), 25–83.]

28. R. Fürth, *Zeitschrift für Physik*, 81, 143–162 (1933).

29. For these investigations, cf. A. Einstein, *Investigations on the Theory of the Brownian Movement*, New York, 1956, which contains all the relevant papers by Einstein and an exhaustive bibliography by R. Fürth. For the experimental work, cf. J. Perrin, *Die Atome* Leipzig, 1920. For the relation between the phenomenological theory and the kinetic theory, cf. also M. v. Smoluchowski, "Experimentell nachwiesbare, der üblichen Thermodynamik widersprechende Molekularphänomene," *Physikalische Zeitschrift*, XIII, 1069 (1912); and K. R. Popper, "Irreversibility, or, Entropy since 1905," *British Journal for the Philosophy of Science*, VIII, 151 (1957). Despite Einstein's epoch-making discoveries and von

Smoluchowski's splendid presentation of their effect (for the latter cf. also *Oeuvres de Marian Smoluchowski*, Cracovie, 1927, Vol. II, pp. 226 ff., 316 ff., 462 ff., and 530 ff.), the present situation in thermodynamics is extremely unclear, especially in view of the continued presence of the ideas of reduction which we criticized in the text above. To be more specific, it is frequently attempted to determine the entropy balance of a complex *statistical* process by reference to the (refuted) *phenomenological* law after which procedure fluctuations are superimposed in a most artificial fashion. For details cf. Popper, *loc. cit.* [For further discussion, see Paul Feyerabend, "On the Possibility of a *Perpetuum Mobile* of the Second Kind," *Mind, Matter, and Method*, ed. P. K. Feyerabend and G. Maxwell (Minneapolis: University of Minnesota Press, 1966), 409–12.]

30. The quantum theory can be adapted to a great many difficulties. It is an open theory in the sense that apparent inadequacies can be accounted for in an *ad hoc* manner, by *adding* suitable operators, or elements in the Hamiltonian, rather than by recasting the whole structure. A refutation of its basic formalism (i.e., of the formalism of quantization, and of noncommuting operators in a Hilbert space or a reasonable extension of it) would therefore demand proof to the effect that *there is no conceivable adjustment of the Hamiltonian, or of the operators used* which makes the theory conform to a given fact. It is clear that such a general statement can only be provided by an *alternative theory* which of course must be detailed enough to allow for independent, and crucial tests.

31. L. Rosenfeld, "Misunderstandings about the Foundations of the Quantum Theory," *loc. cit.*, p. 44.

32. For a very detailed description of a once very influential myth, cf. C. H. Lea, *Materials for a History of Witchcraft*, 3 Vols., New York, 1957, as well as *Malleus Malleficarum*, translated by Montague Summers (who, by the way, counts it "among the most important, wisest [sic!], and weightiest books of the world") London, 1928.

33. Quite clearly, analysis of usage, to take only one example, presupposes certain regularities concerning this usage. The more people differ in their fundamental ideas, the more difficult will it be to uncover such regularities. Hence, analysis of usage will work best in a closed society that is firmly held together by a powerful myth such as was the philosophy in the Oxford of about 10 years ago.

34. Schizophrenics very often hold beliefs which are as rigid, all-pervasive, and unconnected with reality, as are the best dogmatic philosophies. Only such beliefs come to them naturally whereas a professor may sometimes spend his whole life in attempting to find arguments which create a similar state of mind.

35. For details and further references, cf. Section 6 of my "Explanation, Reduction, and Empiricism," *loc. cit.*

36. We assume here that a dynamical rather than a kinematic characterization of motion has been adopted. For a more detailed analysis cf. again the paper referred to in the previous note.

37. For very vivid examples, cf. K. Jaspers, *Allgemeine Psychopathologie*, Berlin, 1959, pp. 75–123.

38. Mae West is by far preferable to the precisionists: "I ain't afraid of pushin' grammar around so long as it sounds good" (*Goodness Had Nothing to Do With It*, New York, 1959, p. 19).
39. I. B. Cohen, Ed., *Isaac Newton's Papers and Letters on Natural Philosophy*, Harvard University Press, Cambridge, Massachusetts, 1958, p. 106.
40. 'Der Gegensatz zwischen der mechanischen und der phaenomenologischen Physik,' *Wärmelehre*, Leipzig, 1896, pp. 362 f.
41. L. Rosenfeld, "Misunderstandings about the Foundations of the Quantum Theory," loc. cit., p. 42.
42. For details concerning the mind-body problem, cf. my "Materialism and the Mind-Body Problem," *Review of Metaphysics*, Sept. 1963.
43. In my paper 'Realism and Instrumentalism' I have tried to show that this is precisely the method which has brought about such spectacular advances of knowledge as the Copernican Revolution, the transition to relativity and to quantum theory.
44. Cf. Section 13 of my "Realism and Instrumentalism," [loc. cit].
45. For support of research the author is indebted to the National Science Foundation and the Minnesota Center for the Philosophy of Science.

Two Concepts of Intertheoretic Reduction

In this paper I reject the widespread view that reductions of scientific theories are all of one basic type. I agree with those who hold that intertheoretic reduction involves the relation of a theory to its special case, but wish to emphasize in how many different ways one theory may constitute a special case of another. In particular, 'special case of *T*' does not mean simply "logical consequence of *T*." I shall argue that we need to recognize at least two main kinds of reduction, which differ both in nature and in scientific function or purpose. "Reduction₁" (as I shall call it) is the achievement of postulational and ontological economy and is obtained chiefly by derivational reduction as described by Nagel; i.e., reduction₁ amounts to the *explanation* of one theory by another. This model is most helpful for understanding what I term "domain-combining" reductions. Reductions of predecessor theories by their successors ("domain-preserving" reductions), on the other hand, normally do not achieve postulational and ontological consolidation, nor are they deductive explanations of the predecessor theories by their successors. They are best described by "inverting" the usual concept of reduction, so that successors are said to *reduce* to their predecessors (not vice versa) under limiting operations and other appropriate transformations. "Reduction₂" (as I label the new concept of reduction) involves a varied collection of intertheoretic relations rather than a single, distinctive logical or mathematical relation. Accordingly, each instance of reduction₂ or partial reduction₂ presents its own technical peculiarities. The great importance of reduction₁ lies in its heuristic and justificatory roles in science. A comparison of the nature and problems of the two kinds of reduction here distinguished will involve us in discussion of several points made in the recent literature of reduction by such writers as Feysabend, Schaffner, and Sklar.

I | Two Concepts of Reduction

Does classical mechanics (CM) reduce to the special theory of relativity (STR), or, on the contrary, does STR reduce to CM? Does Bohr's early quantum theory of the hydrogen atom reduce to a classical theory of the atom in the mathematical limit of high quantum numbers, or, on the contrary, should we say that the classical theory reduces to (is reduced by) Bohr's theory? Surprisingly, the answer is not immediately clear.

Philosophers and scientists often speak of reducing less fundamental theories to more fundamental ones (e.g., physical optics to electromagnetic theory, phenomenological thermodynamics to statistical mechanics). But in the case of CM and STR it is more natural to say that the *more* general STR reduces to the *less* general CM in the limit of low velocities. Epitomizing this intertheoretic reduction is the reduction of the Einsteinian formula for momentum,

$$p = \frac{m_0 v}{\sqrt{1 - v^2/c^2}}$$

(where m_0 is rest mass), to the classical formula, $p = m_0 v$, in the limit as $v \rightarrow 0$. It is quite clear that the STR equation reduces to the CM equation in the limit and that the CM equation does not reduce to the STR equation. Other familiar representatives of a large class of examples of this type are the reduction in the limit of Planck's blackbody-radiation law to the classical Rayleigh formula and the reduction of the van der Waals equation of state to the equation for an ideal gas. This use of 'reduces to' is not only intuitively natural; it is the way physicists and mathematicians, in contrast to most philosophers, usually talk.

We therefore have a striking reversal of terminology between our two sets of examples (reductions involving electromagnetic theory and statistical mechanics on the one hand, and STR, Planck's law, etc., on the other). Since all our examples are reduction paradigms¹ and since we cannot say both that more general or fundamental theories reduce to less general ones *and* that the less general reduce to the more general, in the same sense of 'reduce', I conclude that 'reduce to' means two very different things in these contexts. Our next job is to determine what those meanings are. Since philosophers have preferred to say in all cases that the less general theory reduces to (or is reduced by) the more general, we might suspect that they have neglected an important concept of reduction.²

Reduction is considered by philosophers chiefly to involve an increase in the over-all efficiency of a conceptual scheme, either by outright elimination and replacement of portions that do no work or do their work badly, or (preferably) by consolidation of those parts of the scheme whose

work is discovered to overlap. 'Reduction' means "elimination," "trimming down," "consolidation." This consolidation or general increase in the efficiency of the conceptual scheme may be conceptual—a reduction in the number of independent theoretical assumptions—or ontological, or both at once, as in the reduction of optics to electromagnetic theory. Let us label this kind of reduction "reduction₁."

The reduction of theories by their historical successors constitutes a second great tradition of theoretical reduction, which, however, does not fit the above characterization very well. Reductions in the tradition of Newton and Einstein do not primarily involve the consolidation of concurrent theories, nor are the central issues of these reductions ontological. It is not as if we had both STR and CM and suddenly realized that we did not need CM anymore, or as if we discovered that we do not need to posit both Einsteinian masses and Newtonian masses. Rather, STR came to be entertained specifically as a replacement for CM, and Einstein's theory could not have succeeded without simultaneously making its predecessor fail.

I do not wish to make the distinction between these two reduction traditions seem sharper than it is, but philosophers have too often attempted to fit all instances of reduction into the same mold. It is reductions of the second type in which the heuristic and justificatory functions of reductive relationships are most pronounced. And it is in discussion of these reductions that the concept of reduction now to be introduced becomes especially useful.

The word 'reduction' in ordinary language has many meanings, some of which, like "elimination" or "consolidation," have been exploited by philosophers discussing science. There is another ordinary meaning, or closely related constellation of meanings of 'reduction' which philosophers have not exploited, but which they well might. I refer to the notion of being led back from one thing to another (a sense indicated by the etymology of the word: *reducere*, L.) and to the related notion of transforming something into a different form by performing an operation on it (with an implied continuity of underlying material, perhaps). Some familiar examples, technical and nontechnical are: the "reduction" of ores to their metals by heating with hydrogen; the reduction of wood to pulp by pounding it with a heavy mallet or by grinding it up, as in a paper mill; the fire's reduction of the house to ashes; the reduction of $6xy/2x$ to $3y$ by dividing through by $2x$; and, indeed, the reduction of $m_0v/\sqrt{1 - v^2/c^2}$ to m_0v by taking the limit as v goes to zero. The examples could easily be multiplied. I think a common root image underlies these technical and nontechnical uses of the word 'reduction'.

The exploitation of this common signification in an account of theoretical reduction in science is straightforward. One could say, for example, that STR reduces to CM in the limit of low velocities; i.e., the operation of taking the limit on STR transforms it (leads us back) to a

basic version of CM. I shall term this second kind of reduction "reduction₂" in order to distinguish it from reduction by consolidation (reduction₁). The new concept is useful also in describing other types of intertheoretic relationships such as the syntactical transformations mentioned in sections III and IV below.

Let us note some further differences between reduction₁ and reduction₂. Consolidation of postulates is accomplished by showing that some postulates are logically superfluous, i.e., logically derivative from others. Reduction₁ is therefore essentially *derivational reduction* as analyzed by Nagel in his classic discussion of reduction. Nagel's account will be discussed more fully in the next section. His central requirement is that the reduced theory be a logical consequence of the reducing theory. In other words, such a reduction amounts to a *deductive explanation* of the reduced theory. For obvious reasons, what I have elsewhere termed "domain-combining" reductions are essentially reductions₁.³ The reduction of physical optics to electromagnetic theory was domain-combining in that two theoretically structured domains of phenomena were unified by the reduction. The reduction of one *science* to another would also have to be domain-combining reduction₁. (Actually, domain-combining reductions frequently involve both reduction₁ and reduction₂, as I suggest in section IV.)

By contrast, reduction₂ does not involve the theoretical explanation of one theory by another. Not all reduction is explanation!⁴ Nor are reductions₂ ontological reductions in any real sense. Rather than to effect ontological and conceptual consolidation, the main functions of reduction₂ are justificatory and heuristic. The development of new theoretical ideas is heuristically guided by the requirement that these ideas yield certain established results as a special case (e.g., in the limit), and they are often quickly justified to a degree by showing that they bear a certain relation to a predecessor theory. It was an important confirmation of STR to show that it yielded CM in the correct limit. Paradigms of reduction₂ are the "domain-preserving" reductions of successor theories to their predecessors. The justification derives from the fact that the reduction shows the successor theory to account adequately for the structured domain of phenomena inherited from its successful predecessor; i.e., it is "domain-preserving."⁵ Domain-combining reductions₁ do not justify a new theory in quite the same manner, since, in the domain-combining case, we may have two concurrent theories that are already independently justified. Only the extension of the reducing theory to the new domain need be justified by the reduction. However, reduction *may* also provide a crucial confirmation of a new theory and establish its importance, as in the Maxwell case.

Another point of difference between the two kinds of reduction is that theory T_1 may reduce₂ (or partially reduce₂) to T_2 in several different, noncompeting ways (e.g., as different limits are taken), whereas, strictly

speaking, there can be only one correct way to reduce₁ T_2 to T_1 (since ontological reduction is normally involved).⁶

In the case of reduction₁ the reduced and reducing theories obviously must be logically compatible. Not so for reduction₂. On the contrary, the two theories will *not* be logically compatible where reduction₂ is concerned—unless we recognize reduction₁ as a special case of reduction₂ (see note 6). Also, given its heuristic and justificatory purposes, reduction₂ is not nearly so sensitive to meaning change as is reduction₁. Thus, reduction₂ pretty much avoids the two major objections made to Nagel's treatment of reduction. In fact, reduction₂ provides ways of handling the objections. I now turn to a fuller discussion of these issues.

II | Nagel on Reduction₁

Ernest Nagel's distinction of "homogeneous" and "heterogeneous" reduction somewhat resembles my distinction of domain-preserving and domain-combining reductions and may have suggested it.^{7*} Nagel terms a reduction "homogeneous" when the descriptive vocabulary of the reduced theory T_2 is a subset of the descriptive vocabulary of the reducing theory T_1 . Then T_1 reduces T_2 just in case T_2 is logically derivable from T_1 . The reduction is "heterogeneous" if T_2 contains descriptive terms not in T_1 . In this case straightforward logical derivation is, of course, impossible. Nagel's strategy, in effect, is to turn heterogeneous reduction into homogeneous reduction by conjoining to T_1 a special set H of connective definitions or correlatory laws, which link all the problematic T_2 terms to items in the vocabulary of T_1 . However, it is T_1 (*not* the conjunction of T_1 with H) which is said to reduce T_2 when T_2 is derivable from $T_1 \cdot H$.[†] So, given H , good reasons for believing T_1 are good reasons for believing T_2 . In particular, once the reduction is accomplished, we no longer need believe that T_2 mentions special entities or processes not mentioned by T_1 .

Domain-combining reductions will be heterogeneous in Nagel's sense. Historically distinct domains of phenomena involve different descriptive vocabularies, almost by definition. And domain-combining reductions must involve the logical derivation of one theory from the other (plus connective principles) if it is truly domain-combining, for domain-combining reduction absorbs one theory into another without eliminating the former as incorrect.⁸ This is reduction₁. On the other hand,

* In the book to which Nickles is referring—*The Structure of Science*—Nagel uses the term *heterogeneous* to refer to the reductions that he calls *inhomogeneous* in the reading included in this volume.

† Throughout this reading, Nickles uses a dot to stand for conjunction. Thus, " $T_1 \cdot H$ " should be read as " T_1 and H ."

domain-preserving reductions₂ of a successor theory to a predecessor are homogeneous, so far as terms are concerned, but will almost never be derivational reductions in Nagel's sense. In short, Nagel's analysis of reduction is best regarded as a treatment of domain-combining reduction only. The initial promise of his homogeneous/heterogeneous distinction is lost by Nagel's treatment of all reduction as derivational; in the final analysis he too casts all reduction in essentially the same mold.

Nagel's account is thus left open to two important objections pressed by Feyerabend and by Kuhn.⁹ The first objection is the "meaning change" objection: passing from one theory system to another allegedly involves a conceptual shift so extensive that the same theoretical words do not mean the same thing in the two theories. (For example, 'mass' does not have the same meaning in CM and STR, since mass has very different properties in the two theories.) Therefore the claim that one theory is derivable from the other involves the fallacy of equivocation. It is only an illusion that the two theories contain the same terms.

Why not simply treat these words as distinct terms and introduce bridge laws like *H* to link them?¹⁰ If this could be done, it would turn apparently homogeneous reductions into heterogeneous ones and leave open the possibility that, because of meaning change, no two distinct theories contain the same descriptive vocabulary.

Although this suggestion is useful in particular cases, I believe that as a general strategy it concedes too much to Nagel's critics. Since I have discussed the meaning-change objection more fully elsewhere,¹¹ I shall here point out only that the strategy of treating all homogeneous reductions (and homogeneous parts of reductions) as heterogeneous completely undermines any distinction between the two kinds of reduction which Nagel's terminology may have been intended to reflect. It is to treat the CM-STR reduction as ontological and in essential respects like the reduction of optics to electromagnetic theory. We could no longer distinguish domain-preserving from domain-combining reductions.

I now turn to the second main objection to Nagel's account. This is the objection, stressed by Feyerabend, that reduced and reducing theories are nearly always logically incompatible; therefore one cannot be a logical consequence of the other.¹² At best, it is said, reductions are only "approximative." Although this is true, I think it is important to distinguish those logical incompatibilities which are historical accidents, so to speak, from those based on irreconcilable conceptual or theoretical differences. Except where alternative formulations of essentially the same theory are being investigated, and perhaps even there, it would be good fortune indeed if, without any adjustment of details, one theory should be logically derivable from another theory arrived at by a different line of research. But although the vagaries of historical development moderate our expectation of finding such smooth logical relationships, there is sometimes no conceptual reason why they should not exist. Sometimes compatibility

may be obtained by tinkering with the old theory within the confines of its historical conceptual framework.¹³ And even where such modifications may be too extensive, we can sometimes demonstrate that one theory *partially reduces* the other. Although physical optics is not quite logically compatible with electromagnetic theory, the discrepancies were few and minor enough to handle in one of these ways. Derivational reduction therefore remains a useful concept.

It seems to me that, although their criticisms are legitimate and illuminating up to a point, Nagel's critics tend to forget that philosophical models are deliberately idealized somewhat and are not intended in themselves to give a completely accurate representation of intellectual history. Unfortunately, philosophers almost inveterately forget the limitations of these models and, in time, come to misuse them. We do need to be reminded of their limitations. But it is one thing to remind us of their limitations and another thing to reject them as completely unilluminating. I would go so far as to say that Nagel's derivational reduction is a useful concept even if not one single historical example perfectly exemplifies the pattern he describes. And I would say the same for my distinction between reduction₁ and reduction₂. It is not easy to find pure forms of either type; but is that so surprising? I do not think so—particularly if, as philosophers usually do, we consider only huge, complex theory systems such as CM and quantum theory and neglect theory adjustments and theoretical models on a more manageable scale. Philosophical models are tools that must be applied with care and with attention to the special features of the individual case. It is just the nature of the models themselves to be somewhat idealized.

While rejecting the criticism that *all* claimed reductions₁ and partial reductions₁ fail because of logical incompatibilities of theories, I agree that many of them do. Prominent among the failures are reductions best construed as domain-preserving reductions of successor theories to their predecessors. It is not illuminating to attempt to "correct" CM in order to make it logically compatible with, or partially reducible₁ to STR. But, rather than agree with Nagel's critics that we find no reduction here, I prefer to recognize certain of these important nonderivational intertheoretic relationships as a distinct type of reduction. Reduction₂ is immune to the logical-incompatibility objection. (It is also relatively immune to the meaning-change objection, since its justificatory and heuristic functions do not depend on exact preservation of historical meanings.)

In a broader sense, of course, reduction₂ is also "derivational." "Mathematical derivation," as the phrase is commonly used, includes not only logical deduction but limit processes and approximations of many kinds. In a discussion of Feyerabend's objection, Nagel has recently placed more emphasis on derivation in the wider sense, and, anyway, he has always been aware of some of the theoretical incompatibilities pointed out by his commentators.¹⁴ Nevertheless, Nagel's earlier discussion does treat reduc-

tion as a species of deductive explanation; and in the standard analysis of explanation, to which Nagel subscribes, a deductive explanation is a deductive logical argument in the strict sense.¹⁵ So it is pretty clear that Nagel's original analysis makes intertheoretic reduction derivational in the narrow, not the broad sense.

The existence of reduction, the fact that some paradigm examples of reduction do not fit Nagel's conditions, establishes that Nagel's conditions on reduction are not *necessary* for (all types of) intertheoretic reduction. Two additional counterexamples will establish the *insufficiency* of Nagel's conditions.

First, one may extract from Nagel's discussion a view of the kinetic theory of gases as consisting of three structural components: Newtonian mechanics, assumptions about the nature of matter, and a body of statistical assumptions.¹⁶ Nagel is at least roughly correct in this. Notice, however, that if the kinetic theory contains the laws of CM in some form as a proper part, then it follows from Nagel's account of reduction that kinetic theory reduces CM, CM being trivially derivable from kinetic theory! This result is certainly counterintuitive, since mechanics always has been regarded as the more basic or universal theory, partly for the very reason that it is basic to kinetic theory and to other theories. On Nagel's account, we are not able to distinguish the reduction of a theory from its theoretical application to special domains.¹⁷

A second counterexample, advanced by Lawrence Sklar,¹⁸ is suggested by the Wiedemann-Franz law, according to which the ratio of thermal to electrical conductivity in a metal equals the absolute temperature T multiplied by a constant.¹⁹ But if we employ the biconditional:

The electrical conductivity of a piece of metal is σ
 \equiv the thermal conductivity of the metal is σT (\times a constant)

as a correlatory law linking the vocabulary of a theory of heat conductivity in metals to a theory of electrical conductivity, we could derivationally reduce (or partially reduce) one theory to the other. Though satisfying Nagel's conditions, this would not be a genuine case of reduction.

The problem is that the correlatory statements are in one sense too weak and in another sense too strong. They are too weak in that mere correlation is not enough for reduction. Sklar therefore modifies Nagel's account at least this much: "the place of correlatory laws is taken by empirically established *identifications* of two classes of entities"²⁰ such as light beams and electromagnetic radiation. I shall reserve for later comment this suggestion and a similar one by Kenneth Schaffner.

But in another sense, Nagel's bridge principles may become too strong. In his book he finds it necessary to drop his earlier assumption that the correlatory principles may always be expressed in biconditional form. We may object that, when the correlatory statements become more than

vocabulary translation devices or identifications of entities, it becomes misleading to say that it is T_1 that reduces T_2 rather than the expanded theory $T_1 \cdot H$. Since $T_1 \cdot H$ contains all the terms of T_2 , $T_1 \cdot H$ provides a *homogeneous* reduction of T_2 and therefore fails to establish the superfluity of T_2 concepts. One has not established that the relevant domain can be described and explained by means of T_1 and its vocabulary alone²¹—unless the vocabulary linkages in H have a purely correlatory character. In that case, we could regard H as belonging to the metatheory and regard the original T_1 as reducing T_2 , the reduction being carried out in the metatheory.²² Nagel probably had something like this in mind.

Although these problems in Nagel's account are worthy of more attention than I can give them here, I don't think they show Nagel's account to be fundamentally misconceived.

III | Identificatory Reduction

In this section I shall examine Sklar's and Schaffner's modifications of Nagel's model of reduction. Both writers stress the importance of *theoretical identification* in reduction and urge that Nagel's correlatory laws for heterogeneous reduction are too weak to avoid counterexamples like the Wiedemann-Franz case. Although Sklar and Schaffner are undoubtedly correct that full-fledged theoretical identifications form the core of many scientific reductions, I want to argue, first, that introducing theoretical identification does not completely solve Nagel's problem; second, that identification raises (interesting) new problems of its own; and, finally, that some strong intertheoretic relationships (reductions?) do not involve theoretical identification in the way suggested by these authors. (This last point holds whether or not all reduction requires theoretical identifications.)

First, even if we replace the correlatory biconditionals of Nagel's account with identificatory statements, we may still construct pseudo-reductions like the following:

Light beams refract toward the normal upon transmission into a medium of higher optical density.

Light beams are beams of electromagnetic radiation in frequency range R .

Therefore: Beams of electromagnetic radiation in frequency range R refract toward the normal upon transmission into a medium of higher optical density.

The first premise is a law of geometric optics, the second an identificatory statement; but no one would claim that we have hereby reduced a law of

electromagnetic theory to a law of geometric optics, or that we have partially reduced electromagnetic theory to geometric optics! We have here a general problem for the standard account of scientific explanation.²³ Nagel's problem is just a special case of the explanation problem, since *derivational* reduction is regarded as the explanation of the reduced theory by the reducing theory.

Second, Sklar's identificatory requirement is not as specific as it might be. Does Sklar intend that *all* vocabulary bridge principles in heterogeneous reduction be theoretical identifications? This would be a very strong requirement and would raise problems (of which he is already aware) concerning the identification of properties. Or is he merely claiming that all genuine reductions are basically ontological?—which is false. Perhaps, like Schaffner, he would confine identifications to the entities in the domains of individuals of the two theories in question. However, this measure in itself is insufficient to rule out even the Wiedemann-Franz example, for both theories concern exactly the same entities (electrons, etc.). Still, there is a sense in which the two theories are not about the same things.

In any event, it is frequently very difficult to know what entities, states, or processes to identify with which. This problem becomes especially thorny when one attempts to reduce nonstatistical theories like CM or phenomenological thermodynamics to statistical theories like QT or statistical mechanics. We should like to identify the thermodynamical property of temperature with the statistical property of mean kinetic energy of the molecules; otherwise we would be hard pressed, after the reduction, to say that the two theories were really about the same thing. (In short, we would have no reduction!) But, as Sklar himself has pointed out, there usually are alternative concepts of mean kinetic energy and other key terms available in the statistical theory, based on different ways of averaging, for example (ensemble averages, time averages, etc.). Although the various values of mean kinetic energy are virtually the same, practically speaking, the theoretical concepts are not the same. The problem then is which statistical concept to identify with temperature. Distinct alternatives offer different advantages: we want our new concepts of heat, temperature, entropy, etc., to act as much like the old concepts as possible, but alternative identifications bring out different patterns of resemblance between the two theories. Any way we do it, the new concepts behave like the old in some ways but not in others, which ways depending on the prior identification.

I now want to point out that, in cases like that suggested by the relation of CM to quantum theory (QT), we may deliberately give up the most natural ontological and theoretical identifications at one point in order to obtain a significant correlation elsewhere. Particularly in the reduction of theories to their successors, ontological issues are rarely the

main concern. Scientists are here far more interested in the various formal and conceptual linkages of the theories. (These are prior to ontological questions anyway, at least in this range of cases.)

Thus, it would seem more natural to connect the classical concept of particle position with the quantum-mechanical concept of position, which, though different, is far more like the classical concept than is the QT concept of the *expectation value* of position, a concept involving a statistical average over an ensemble of identically prepared systems.²⁴ Yet when we compare the structure of the two theories, it proves far more illuminating to correlate classical values with quantum expectation values. The result is expressed in the well-known Ehrenfest's theorem, which states, roughly, that any relation that appears in CM must be valid as a relation between QT expectation values. Ehrenfest's theorem helped to justify the new QT for physical reasons as well as formal. As Jammer points out,

Ehrenfest's affirmation of Newton's second law in the sense of averages taken over the wave packet had a great appeal to many physicists and did much to further the acceptance of the theory.²⁵

The point is, in these contexts in which justification and heuristics are the main concerns, we deliberately forswear attempting the ontological and conceptual identification which is so essential to other forms of reduction (though physicists did misleadingly come to think of the wave packet as a physical surrogate for the individual particle).

Now one might argue that for this very reason the QT type of case should not be considered a genuine instance of reduction; and likewise for the "reduction" of CM by STR and the other troublesome counterexamples to derivational reduction.

In reply, let it be clear that I am not claiming that the CM-QT case, which I just used to make a point, is an instance of reduction, though I think the relationship sketched might be termed "reductive" in a wider sense. Such relationships merit close philosophical study; it is conceivable that *some* syntactical transformations be recognized as full-fledged reductions. At any rate, having given up the ideal of the logical consequence relation as *the* reduction relation, we are better prepared to handle Feyerabend's objection to Nagel's account that reduction in science is nearly always approximative. Reduction is rarely a matter of simple logical relationships.

IV | Approximative Reduction

Both Sklar and Schaffner are cognizant of Feyerabend's objection. To handle approximative reduction, and perhaps meaning change as well,

Schaffner requires not that the historical, reduced theory T_2 (the "secondary theory") but a "corrected secondary theory" T_2^* be logically derivable from the reducing theory T_1 (the "primary theory").²⁶ His final three conditions express the relation of T_2^* to T_2 :

(3) T_2^* corrects T_2 in the sense of providing more accurate experimentally verifiable predictions than T_2 in almost all cases . . . , and should also indicate why T_2 was incorrect (e.g., crucial variable ignored), and why it worked as well as it did. (4) T_2 should be explicable by T_1 in the nonformal sense that T_1 yields a deductive consequence (when supplemented by . . . [identificatory statements]) T_2^* which bears a close similarity to T_2 and produces numerical predictions which are "very close" to T_2 's. Finally (5) the relations between T_2 and T_2^* should be one of strong analogy (144).²⁷

Schaffner's proposal has the merit of making reduction much more than a matter of logical derivation. The relation of T_1 to T_2 is not merely the logical-consequence relation but a "product" of that with the analogy relation. But while most suggestive, the importance of the analogy relation of T_2 and T_2^* in Schaffner's analysis is in another respect unfortunate in that (as he himself points out) the concept of analogy has itself received no satisfactory philosophical analysis. Here the concept of reduction₂ may be of some help. Presumably, limit relationships and the like will sometimes be useful in spelling out the analogy of T_2 and T_2^* in particular cases. That is, in cases of approximative reduction, we can regard T_2^* as a (fictitious) successor theory to the historical T_2 , and in many of these cases reduction₂ techniques should be useful in spelling out the analogy. There is no reason to think that reductions₂ can in all cases take over the work of Schaffner's analogy relation, but when it can we shall have T_1 reducing, T_2 approximatively by reducing, T_2^* , which in turn reduces₂ to T_2 .

Schaffner is sensitive to the fact that general models of reduction cannot have built into them a set of detailed conditions sufficient to handle the application of the models to all relevant historical cases. He is right not to attempt a general analysis of his analogy relation. What he does not see is that, although most historical reductions are "approximative" in Feyerabend's wide sense, they are not all approximative reductions of the same kind. My present criticism is that Schaffner's interesting proposal still is made within the Nagel framework of one general type or model of reduction, with the important additional idea of varied conditions of application. Schaffner's model is still basically derivational.

Nagel, his commentators, and his critics alike all make the mistake of regarding reduction₂ as an imperfect attempt at derivational reduction₁, as a failure to achieve reduction₁. But this view misses completely the point and nature of reduction₂. It is misleading to label reduction₂ in itself "approximative" at all; rather, just as we distinguish perfect from approximative reduction₁, we must distinguish perfect from approximative

reduction₇. Reduction₂ can sometimes help us analyze approximative reduction₁, as indicated, but it can do so only by being a separate kind of reduction. In sum, it is simply false that any reductive relation of logically incompatible theories *must* be approximative (in the sense of failing to fit any general model of reduction).

Although he has made a start, I do not believe that Schaffner has gone far enough in de-emphasizing the importance of logical derivation in many reductions. Often not even the corrected secondary theory T_2^* will be derivable from T_1 ; rather, we get T_2^* by applying limit operations or other transformations to the law statements of T_1 . We do not get a corrected secondary theory analogous to the historical CM by logical derivation from STR. It is only by taking the limit of certain STR equations as the velocity of the system goes to zero (or as the maximum signal velocity goes to infinity) that we get the post-Einstein version of CM.²⁸

Another type of case in which derivation is conspicuous by its absence is suggested by Ehrenfest's theorem, mentioned in section III. In this type of case, one theory has equations of the form $a = b$, $c = d$, etc., while its successor has equations of the form $\langle a \rangle = \langle b \rangle$, $\langle c \rangle = \langle d \rangle$, etc. That is, the relations of the variables of the first theory are valid as relations of the expectation values of those variables in the second theory.²⁹ (Compare the construction of QT from CM equations by replacing Poisson brackets by commutators.) What sort of intertheoretic relation is this?

The physicist H. A. Kramers terms the similar relation expressed by the "generalized Ehrenfest theorem" one of "formal analogy," and most writers would surely agree.³⁰ However, philosophers like C. G. Hempel and Mary B. Hesse have also used a more restricted concept of "formal analogy" or model. Hesse speaks of formal analogy as "the one-to-one correspondence between different interpretations of the same formal theory"³¹ (as in the case of acoustics and electric circuit theory); similarly, for Hempel two theories are "syntactically isomorphic" only if they are based on *identical* syntactical calculi.³² In our case, however, we are not concerned with theories having the *same* formal calculus: the relation of the formal calculi is part of the problem! Surely the two theories *are* formally isomorphic or formally analogous in a wider sense. We can see that one of our idealized theories is a transformation of the other—we pass from the second theory to the first by the operation of replacing expectation values of variables by the variables themselves.³³ This more general notion of sameness of form would seem to be of great importance in twentieth-century physics, with its emphasis on invariance requirements of various sorts. Whether or not this important type of relationship is reductive, it fits a now familiar pattern—that of reduction₇.

V | Reduction₂ Again

The following schema suggests itself. Instead of a single reduction relation (the logical-consequence relation), we now recognize a set of intertheoretic operations O_i (or corresponding relations) which are reductive under certain conditions. Then we might write $O_2[O_1(T_1)] \rightarrow T_2$ or simply $O_2O_1(T_1) \rightarrow T_2$, signifying that by performing operations O_1 and O_2 on theory T_1 we get T_2 . No attempt will be made to build this notational scheme into a general account of reduction₂. It is important to keep in mind that only very rarely will *all* the equations of T_2 reduce₂ to equations of T_1 under the operations. ' T_1 ' and ' T_2 ' usually will stand for theory parts.³⁴ Nonetheless, partial reduction of this sort remains an indispensable conceptual tool in theoretical research.

To map one theory into or onto another reductively, do we sometimes need to perform operations on both theories, or can we always compound operations on the successor theory? Compounding is possible in this way only if the inverse of the operation on the predecessor theory exists, and this generally will not be the case. What, for example, is the inverse of a limit operation? It is conceivable that two different relativity theories be isomorphic only in the limit of low velocities *for both*. Using the inverse operation obviously would lead to error. Notice, however, that cases in which operations must be performed also on the reduced theory T_2 depart somewhat from the image underlying reduction₂. Although their relation is interesting and may be theoretically important, we probably would not say the two relativity theories reduced to each other at all. (They each reduce to a third theory.) Since inverses of syntactical transformations do exist, either of two theories so related may be reduced₂ to the other. So reduction₂ relations may be either symmetric or asymmetric, it would appear.

Given two physical theories or theory parts (A) and (B), in what scientifically permissible and interesting ways can we go from one to the other so that the relation is reductive? Logical derivation, if possible, is certainly legitimate. Other devices are more problematic. Mathematical differentiation and integration are interesting possibilities, perhaps, and syntactic transformations and limits are devices previously mentioned. Frequently, sums are replaced by integrals, as in the transition from quantized to classical results. At several points in physics discrete and continuous treatments of the phenomena are interestingly related. I must restrict the discussion here to raising some questions about simple limit relationships and ignore the technical realities faced by the mathematical physicist.

Let us consider a particular, fictional example, which, for simplicity, concerns two little theories or theory parts of one sentence each:

$$(A) w = ax + 2y + g \quad (B) z = bx + ey + d$$

where w, x, y, z represent "physical variables" and a, b, d, e, g are numerical constants. First assume that $a = b, e = 2$. Then the two equations differ only in the addends g and d . Suppose (A) is or belongs to the old theory, with $g = 0$, and (B) is a surprising new theory for which $0 \neq d \ll 1$. Is it legitimate to say that (B) reduces₂ to (A) in the limit as $d \rightarrow 0$? If so, why is it not equally legitimate to say that (A) reduces₂ to (B) in the limit as $g \rightarrow d$? Note that g and d are physical *constants*.

Two standard ways of reducing₂ STR to CM are usually cited: take the limit of low velocities or allow the maximum signal velocity c to approach infinity. Similarly, it is sometimes claimed that the old QT partially reduces to CM either in the limit of high quantum numbers or as Planck's constant $h \rightarrow 0$. It is important to notice the difference in the two limit procedures in each case. The first of each pair is an ordinary case of letting a variable approach a limit, but the second permits a *constant* to act as a variable approaching a limit. If we are going to permit constants of nature, i.e., numerical constants, to approach new values—and physicists (including Planck and Bohr³⁵) sometimes do this with Planck's constant and the velocity of light—there are still more interesting possibilities. Suppose that $e = 0$ in the new theory. Can (B) reduce₂ to (A) by our letting e now approach 2? Is this any more or less justifiable than letting $y \rightarrow 0$ in the old theory, thus reducing₂ (A) to (B)? Why in this case should we say that (A) reduces₂ to (B), and not vice versa?—Simply because the limiting operation was performed on (A)?

One intuitively feels that if one theory becomes another when a certain limit is taken, the former is somehow a generalization of the latter, and hence limit reductions, unlike syntactical transformations, are asymmetric: the general case reduces₂ to the special case. But the asymmetry is not preserved in our case, for we previously carried through the reduction in the opposite direction; i.e., it seems that each theory can be considered a generalization of the other on exactly the same grounds! And if constants can be made to approach a different value, why not the coefficient 2 in (A)? By letting $2 \rightarrow 0$ we can also eliminate the y term. And by this means every equation reduces to every other—a complete trivialization of the concept of intertheoretic reduction. Any physical-constant coefficient can be eliminated by taking it to 1 (take additive factors to zero). Any expression whatever may be introduced or eliminated from an equation by these means. Clearly we must say that letting numerical constants change value is mathematically illegitimate.

The same trivialization obviously follows also from unrestricted use of syntactic transformations. Under what conditions do such transformations and limit taking amount to genuine reductions?

I do not believe there is a general formula for reductive relationships any more than there is an interesting common structure to all theories: these are places where we must look and see—we must examine each case on its individual merits. Nevertheless, there are some informal observations

we can make. For one thing, reductively related theories must concern the same or significantly overlapping domains of phenomena. For another thing, when one whole set of equations is to be reduced to another entire set, complete trivializations like the above are harder to come by because of the presence of the expressions in several members of the sets. Indeed the triviality of our example was exaggerated. If g is a variable on the real numbers while $d = 2$, it is possible that (A) be a generalization of (B); but it is then impossible that (B) also be a generalization of (A), no matter how the other variables and factors are related. For even if we permitted constants to be assigned new values (e.g., limiting values), there is no way of making constant d serve as a variable. Precisely this asymmetry obtains between the generalized Balmer law for spectral wave lengths λ , $1/\lambda = R(1/m^2 - 1/n^2)$, and the Balmer law proper, $1/\lambda = R(1/2^2 - 1/n^2)$, where R is Rydberg's constant and m, n assume integral values (with some restrictions). The Balmer law is clearly the special case of the generalized law when $m = 2$. But there is no sense in which the general law is a special case of the Balmer law. Note, incidentally, that the Balmer law is a special case without being a *limiting* case of the general law. (It would be misleading to say that the Balmer law emerges as m approaches 2 as a limit.) This suggests there may be several distinct general-special case relationships worth exploring.

Another consideration that will weed out many unwanted "reductions" is that we usually entertain reduction claims concerning only the reduction₂ of theories to previously established theories—to respected predecessors. $O_1(T_1) \rightarrow T_2$ is of no importance unless T_2 is of some interest. Successful predecessor theories impose constraints on the physical variables and their mathematical relations in potential successor theories in the sense that any discrepancy must be accounted for. (The predecessor's usual general agreement with the facts is only one reason for this.) A gas theory or theory of solids that did not contain the temperature variable could not be taken seriously, whether or not it could be easily transformed into an existing theory. And energy radiated from a body had better depend on the fourth power of the temperature of the body.

Not that a theory must be entirely successful to be a worthy precedent: it may be a mere "model," deliberately oversimplified—the minimal realization of a promising theory program. A good example is just the early statistical-mechanical models of solids and gases and their later sophistication by the successive introduction of new degrees of freedom which would "freeze up" at lower temperatures and "unfreeze" at higher temperatures. (See note 2.)

Not surprisingly, even a predecessor theory known to be defective largely determines the range of admissible new proposals. This respect for tradition is, of course, more than blind devotion to the past, but does represent an element of conservatism in scientific research. Radical new proposals are frequently not taken seriously, and, if eventually they are, it

is only at the cost of regarding the predecessor theory as less successful than we previously believed.

Finally, it is important that the reductive operations make physical sense. By taking the limit as temperature T goes to infinity, we may be able to eliminate T or a more complex factor involving T from some theory. But an infinite temperature does not make physical sense. Nor does it make physical sense to allow T to vary normally but to eliminate it by introducing a special new multiplicative factor to which no physical interpretation can be assigned and on which the limits are taken.³⁶

Indeed, it makes no physical sense to permit physical constants to vary ($h \rightarrow 0$, $c \rightarrow \infty$). But the reason it is illuminating in such cases as *these* is that these values were in a sense tacitly assumed in the predecessor theory or at least were compatible with it. Thus we can think of h and c as variables in a noncommittal metalanguage.³⁷ The moral here is that the illumination that highlighting a certain relationship brings will depend largely upon the internal makeup of the theories concerned, upon problems to which they happen to be applied, and sometimes upon historical factors. One reason that tinkering with Planck's constant is more interesting and more significant than tinkering with other physical constants (and similarly for the velocity of light in STR) is surely that it is so fundamental to quantum theories that it practically has come to characterize QT statements. The presence of Planck's constant is one common way of individuating classical from QT ideas.

The requirements that there be a fruitful predecessor theory and that the limit operations make physical sense as well as mathematical sense do much to solve the problem of trivialization for reduction.² The problem of finding reductive relations that are weak enough to be exemplified in real science and yet strong enough to be interesting is nothing new after all. Even derivational reduction faces its own peculiar form of the problem of trivialization,* as Nagel points out in his section on "Nonformal Conditions for Reduction."^{38, 39}

■ | Notes

1. Lack of space forces me to rely more heavily than I would wish on standard "textbook" examples of reduction, which are notoriously controversial.

* The problem of trivialization for derivational reduction is the problem of distinguishing trivial from important scientific achievements. Mere logical deducibility from some set of premises is not sufficient to render a reduction nontrivial. Epistemic conditions must also be satisfied. Among these are Nagel's requirements that the reducing theory be empirically well confirmed and that it integrate a wide range of laws that had previously been accepted on independent evidential grounds.

2. With few exceptions philosophers have attempted to force all scientific reductions into the same mold, the mold varying somewhat from philosopher to philosopher. This uniformity of treatment is somewhat surprising in view of the fact that reduction takes place at different levels, each with its special problems, and that reduction fulfills a variety of scientific and philosophical functions. Reduction sometimes concerns the relation of entire scientific disciplines, sometimes theories within a discipline, sometimes laws and models within a single-theory framework (e.g., the reduction of a model of solids by a generalized model involving additional degrees of freedom, and even the reduction of Balmer's formula for spectral frequencies by the generalized Balmer formula), and sometimes terms, in a sense (e.g., the ontological reduction of light to electromagnetic radiation or of enzymes to polypeptide chain structures). An example of a problem that is serious at one level but not at others is what I have elsewhere termed "cross-category" problems of explanatory failure which threaten to obstruct the reduction of some disciplines to others. See my ["Theory Generalization, Problem Reduction, and the Unity of Science," *PSA 1974, Boston Studies in the Philosophy of Science*, vol. 32, ed. A. Michalos and R. S. Cohen (Dordrecht, Netherlands: Reidel, 1976), 31-74,] for discussion. See also Ernest Nagel, *The Structure of Science* (New York: Harcourt, Brace, & World, 1961), pp. 399ff.

3. The distinction between domain-combining and domain-preserving reductions was introduced in my "Heuristics and Justification in Scientific Research," § 3, in Frederick Suppe, ed., *The Structure of Scientific Theories* (Urbana: Univ. of Illinois Press, 1973).

4. Approximative reductions, including reduction₂, might be said to explain why the predecessor theory worked as well as it did, as several writers have pointed out.

5. Predecessor theories are rarely completely successful, and the successor is, of course, required to be domain-preserving only where the predecessor was successful. A good example of domain-preserving reduction is the work on statistical-mechanical models of solids stemming from Einstein's simple lattice model of 1907 and similar work on models of gases. In order to account for serious deviations from the predicted specific heats at temperature extremes, the models were generalized by introducing new degrees of freedom, the thermodynamic effects of which became significant only under relatively extreme conditions. Thus the reducibility of the generalized models to the earlier, partially successful models provided an easy preliminary confirmation *en bloc* of the new models, on the side of justification. And on the side of discovery, the belief that such reducibility in the limit must obtain, provided an obvious but powerful heuristic and plausibility check for the development of new theoretical models by "rational generalization" of the old. (On the function of reduction as confirmation, see Lawrence Sklar, *Intertheoretical Reduction in Natural Science*, unpublished doctoral dissertation, Princeton, 1964.)

6. It is possible for a theory to reduce₁ one theory and to reduce₂ to another theory under appropriate operations. Indeed, we should expect that the relation of a theory to its highly successful but defective predecessor will be importantly different from its relation to another, concurrent theory which it absorbs. In short, a theory may, by virtue of its reductions, be both domain-combining and domain-preserving. Perhaps it is just because many domain-combining theories are at least

partially domain-preserving that philosophers have assimilated the two kinds of reduction.

The two models of reduction must be applied to historical examples with great care. I am not advancing the simplistic view that all or even most historical reductions fit neatly into one or the other of the schemata. For one thing, many of the reductions will be partial; for another, reduction₁ is rarely strictly derivational, as Nagel originally seemed to think, but usually involves limits or other special case restrictions on some of the variables of T_1 —initial and boundary conditions for the reduction. This suggests that we might construe reduction₁ as a special case of reduction₂ in which the chief transforming operation is “take the logical consequences.” Still, we must distinguish those cases in which the restrictions apply to T_1 variables that do not belong also to T_2 (in which case the two theories may still be logically compatible) from those cases like STR-CM in which the variable restricted in T_1 (velocity) also belongs to T_2 .

7. Nagel, *op. cit.*, ch. 11. Nagel seems to have in mind the same two reduction traditions that my distinction is intended to reflect.

8. Nagel apparently overlooks the fact that the correlatory sentences H will include initial and boundary conditions (restrictions on T_1) as well as connective definitions and laws. Cf. Nagel, *op. cit.*, p. 34, and fn 6 above.

9. P. K. Feyerabend, “Explanation, Reduction, and Empiricism,” in H. Feigl and G. Maxwell, eds., *Minnesota Studies in the Philosophy of Science*, vol. III (Minneapolis: Univ. of Minnesota Press, 1962). T. S. Kuhn, *The Structure of Scientific Revolutions* (Chicago: University Press, 1962), pp. 100ff.

10. This interesting suggestion has been advanced by Morgenbesser.

11. [“Reduktion/Reduktionismus,” in *Handbuch Wissenschafts theoretischer Begriffe*, ed. Josef Speck et al. (Goettingen: Vandenhoeck and Ruprecht, 1980), 548–53.]

12. Several writers have pointed out that this second objection is inconsistent with the first: “incommensurable” theories cannot be logically incompatible.

13. Morgenbesser and Schaffner have both made this point. Schaffner speaks of “the corrected secondary theory” (see sec. IV below for discussion).

14. See Nagel’s “Issues in the Logic of Reductive Explanations” in H. Kiefer and M. Munitz, eds., *Mind, Science and History* (Albany: SUNY Press, 1970). Nagel points out the logical incompatibility of Kepler’s and Galileo’s theories with Newton’s in *The Structure of Science*, p. 58.

15. Whether deductive explanation is itself better regarded as derivational in the wider sense, and whether the same is true of the view that theories are deductive logical systems, are tales for another time.

16. Nagel writes: “The theory of gases retains the fundamental assumptions of Newtonian mechanics,” “. . . supplemented by a hypothesis about the molecular constitution of a gas” (he adds in a related passage), plus various statistical assumptions (*The Structure of Science*, pp. 345ff).

17. The easiest way out of this puzzle is probably to modify slightly this conception of the gross structure of kinetic theory. That Nagel’s account of derivational reduction is in need of adjustment was pointed out in fns 6 and 8 above. Also, the

puzzle nicely brings out the interconnection of the problems of the structure of scientific theories and intertheoretic reduction.

18. "Types of Inter-Theoretic Reduction," *British Journal for the Philosophy of Science*, XVIII, 2 (August 1967): 109–124. As his title indicates, Sklar is one philosopher interested in distinguishing types of reduction.

19. See, for example, Arnold Sommerfeld, *Thermodynamics and Statistical Mechanics* (New York: Academic Press, 1964 edition), p. 276.

20. Sklar, *op. cit.*, p. 120.

21. On this point see C. G. Hempel, "Reduction: Linguistic and Ontological Issues," in S. Morgenbesser *et al.*, eds., *Philosophy, Science and Method: Essays in Honor of Ernest Nagel* (New York: St. Martins, 1969), p. 189; Sklar, *op. cit.*, p. 119; and Nils Røss-Hansen, "The Reduction of Biology to Physical Science," *Synthese*, xx, 2 (August 1969): 281.

22. The theory/metatheory distinction here is context-dependent and is not intended to demarcate essentially different modes of scientific inquiry. In particular, *H* will consist of empirically supported theoretical statements.

23. See my "Covering Law Explanation," *Philosophy of Science*, [38] (1971): 542–561, and ["Theory Generalization, Problem Reduction, and the Unity of Science,"] *op. cit.*

24. Similarly, I am more like a particular eighteenth-century man than like the statistician's "average man."

25. Max Jammer, *The Conceptual Development of Quantum Mechanics* (New York: McGraw-Hill, 1966), pp. 363f.

26. Kenneth Schaffner, "Approaches to Reduction," *Philosophy of Science*, [34] (1967): 137–147.

27. Everyone recognizes that it is no final analysis to say that the reducing theory explains why the reduced theory works where it does. What we are looking for is the mechanism involved.

28. The most that we can strictly *derive* from STR is *conditional* statements like "If system velocity is negligible relative to the velocity of light, *then*. . ."

29. The symbolism here leaves it unclear whether the expectation values of individual variables or of some combination of variables is being employed; but for present purposes it doesn't matter.

30. H. A. Kramers, *Quantum Mechanics* (New York: Dover, 1964), p. 111.

31. M. B. Hesse, *Models and Analogies in Science* (London: Sheed & Ward, 1963), p. 75.

32. C. G. Hempel, *Aspects of Scientific Explanation* (New York: Free Press, 1965), p. 436.

33. I am not against saying that this sort of relationship amounts to modeling the successor theory on the predecessor, although it differs from other cases of modeling in which the two theories have the same formalism but concern entirely different domains of phenomena. But the very diversity of kinds of models makes it relatively unilluminating to introduce model talk at this point.

34. Alternatively, the arrow can be interpreted as mapping *into* T_2 . Actually, we should distinguish two senses in which a reduction₃ can be complete: when every sentence of T_1 is mapped to a (any) sentence of T_2 and when every sentence of T_2 has a sentence of T_1 mapped to it. There are two corresponding ways in which a reduction may be incomplete or partial.

35. For example, Jammer, *op. cit.*, p. 109, quotes Planck as saying in his *Vorlesungen über die Theorie der Wärmestrahlung*, "The classical theory can simply be characterized by the fact that the quantum of action becomes infinitesimally small."

36. Although "physical sense" is an important consideration, idealization, sometimes to the point of inconsistency (e.g., point masses, point particles), plays a central role in theoretical research. For a thorough discussion of this point, see Dudley Shapere, "Notes toward a Post-Positivist Interpretation of Science," in P. Achinstein and S. Barker, eds., *The Legacy of Logical Positivism* (Baltimore: Johns Hopkins Press, 1969).

37. For further discussion of limit relationships, the reader is referred to M. Strauss, "Intertheory Relations," in P. Weingartner and G. Zecha, eds., *Induction, Physics, and Ethics* (Dordrecht: D. Reidel, 1970), which became known to me after this paper was written. Strauss points out that when constants are taken to limits, they are being regarded as variables in the metatheory. This must be correct, but, of course, does not in itself explain our reluctance to treat other constants in the same way.

38. Nagel, *The Structure of Science*, p. 358.

39. I wish to thank Professors C. G. Hempel and Richard Grandy of Princeton University and my colleagues, Professors F. L. Will and Frederick Suppe, for helpful comments on earlier drafts.

1953 and All That: A Tale of Two Sciences

"Must we geneticists become bacteriologists, physiological chemists and physicists, simultaneously with being zoologists and botanists? Let us hope so."

—H. J. Muller, 1922¹

1 | The Problem

Toward the end of their paper announcing the molecular structure of DNA, James Watson and Francis Crick remark, somewhat laconically, that their proposed structure might illuminate some central questions of genetics.² Thirty years have passed since Watson and Crick published their famous discovery. Molecular biology has indeed transformed our understanding of heredity. The recognition of the structure of DNA, the understanding of gene replication, transcription and translation, the cracking of the genetic code, the study of gene regulation, these and other breakthroughs have combined to answer many of the questions that baffled classical geneticists. Muller's hope—expressed in the early days of classical genetics—has been amply fulfilled.

Yet the success of molecular biology and the transformation of classical genetics into molecular genetics bequeath a philosophical problem. There are two recent theories which have addressed the phenomena of heredity. One, *classical genetics*, stemming from the studies of T. H. Morgan, his colleagues and students, is the successful outgrowth of the Mendelian theory of heredity rediscovered at the beginning of this century. The other, *molecular genetics*, descends from the work of Watson and Crick. What is the relationship between these two theories? How does the molecular theory illuminate the classical theory? How exactly has Muller's hope been fulfilled?

There used to be a popular philosophical answer to the problem posed in these three connected questions: classical genetics has been reduced to molecular genetics. Philosophers of biology inherited the notion of reduction from general discussions in philosophy of science, discussions which usually center on examples from physics. Unfortunately attempts to apply

this notion in the case of genetics have been vulnerable to cogent criticism. Even after considerable tinkering with the concept of reduction, one cannot claim that classical genetics has been (or is being) reduced to molecular genetics.³ However, the antireductionist point is typically negative.⁴ It denies the adequacy of a particular solution to the problem of characterizing the relation between classical genetics and molecular genetics. It does not offer an alternative solution.

My aim in this paper is to offer a different perspective on intertheoretic relations. The plan is to invert the usual strategy. Instead of trying to force the case of genetics into a mold, which is alleged to capture important features of examples in physics, or resting content with denying that the material can be forced, I shall try to arrive at a view of the theories involved and the relations between them that will account for the almost universal idea that molecular biology has done something important for classical genetics. In so doing, I hope to shed some light on the general questions of the structure of scientific theories and the relations which may hold between successive theories. Since my positive account presupposes that something is wrong with the reductionist treatment of the case of genetics, I shall begin with a diagnosis of the foibles of reductionism.

2 | What's Wrong with Reductionism?

Ernest Nagel's classic treatment of reduction⁵ can be simplified for our purposes. Scientific theories are regarded as sets of statements.⁶ To reduce a theory T_2 to a theory T_1 , is to deduce the statements of T_2 from the statements of T_1 . If there are nonlogical expressions which appear in the statements of T_2 , but do not appear in the statements of T_1 , then we are allowed to supplement the statements of T_1 with some extra premises connecting the vocabulary of T_1 with the distinctive vocabulary of T_2 (so-called *bridge principles*). Intertheoretic reduction is taken to be important because the statements which are deduced from the reducing theory are supposed to be explained by this deduction.

Yet, as everyone who has struggled with the paradigm cases from physics knows all too well, the reductions of Galileo's law to Newtonian mechanics and of the ideal gas laws to the kinetic theory do not exactly fit Nagel's model. Study of these examples suggests that, to reduce a theory T_2 to a theory T_1 , it suffices to deduce the laws of T_2 from a suitably modified version of T_1 , possibly augmented with appropriate extra premises.⁷ Plainly, this sufficient condition is dangerously vague.⁸ I shall tolerate its vagueness, proposing that we understand the issue of reduction in genetics by using the examples from physics as paradigms of what "suitable modifications" and "appropriate extra premises" are like. Reductionists claim that the relation between classical genetics and molecular biology

is sufficiently similar to the intertheoretical relations exemplified in the examples from physics to count as the same type of thing; to wit, as intertheoretical reduction.

It may seem that the reductionist thesis has now become so amorphous that it will be immune to refutation. But this is incorrect. Even when we have amended the classical model of reduction so that it can accommodate the examples that originally motivated it, the reductionist claim about genetics requires us to accept three theses:

- R1 Classical genetics contains general laws about the transmission of genes which can serve as the conclusions of reductive derivations.
- R2 The distinctive vocabulary of classical genetics (predicates like '① is a gene', '① is dominant with respect to ②') can be linked to the vocabulary of molecular biology by bridge principles.
- R3 A derivation of general principles about the transmission of genes from principles of molecular biology would explain why the laws of gene transmission hold (to the extent that they do).

I shall argue that each of the theses is false, offering this as my diagnosis of the ills of reductionism.

Before offering my criticisms, it may help to explain why reductionism presupposes (R1)–(R3). If the relation between classical genetics and molecular biology is to be like that between the theory of ideal gases and the kinetic theory (say), then we are going to need to find general principles, identifiable as the central laws of classical genetics, which can serve as the conclusions of reductive derivations. (We need counterparts for the Boyle-Charles law.) These will be general principles about genes, and, because classical genetics seems to be a theory about the inheritance of characteristics, the only likely candidates are laws describing the transmission of genes between generations. [So reductionism leads to (R1).] If we are to derive such laws from molecular biology, then there must be bridge principles connecting the distinctive vocabulary figuring in the laws of gene transmission (presumably expressions like '① is a gene', and perhaps '① is dominant with respect to ②') with the vocabulary of molecular biology. [Hence (R2).] Finally, if the derivations are to achieve the goal of intertheoretical reduction then they must explain the laws of gene transmission. [(R3).]

Philosophers often identify theories as small sets of general laws. However, in the case of classical genetics, the identification is difficult and those who debate the reducibility of classical genetics to molecular biology often proceed differently. David Hull uses a characterization drawn from Dobzhansky: classical genetics is "concerned with gene differences; the operation employed to discover a gene is hybridization: parents differing in some trait are crossed and the distribution of the trait in hybrid progeny

is observed.”⁹ This is not unusual in discussions of reduction in genetics. It is much easier to identify classical genetics by referring to the subject matter and to the methods of investigation, than it is to provide a few sentences that encapsulate the content of the theory.

Why is this? Because when we read the major papers of the great classical geneticists or when we read the textbooks in which their work is summarized, we find it hard to pick out *any* laws about genes. These documents are full of informative statements. Together, they tell us an enormous amount about the chromosomal arrangement of particular genes in particular organisms, about the effect on the phenotype of various mutations, about frequencies of recombination, and so forth.¹⁰ In some cases, we might explain the absence of formulations of general laws about genes (and even of reference to such laws) by suggesting that these things are common knowledge. Yet that hardly accounts for the nature of the textbooks or of the papers that forged the tools of classical genetics.

If we look back to the pre-Morgan era, we do find two general statements about genes, namely Mendel's Laws (or “Rules”). Mendel's second law states that, in a diploid organism which produces haploid gametes, genes at different loci will be transmitted independently; so, for example, if *A*, *a* and *B*, *b* are pairs of alleles at different loci, and if an organism is heterozygous at both loci, then the probabilities that a gamete will receive any of the four possible genetic combinations, *AB*, *Ab*, *aB*, *ab*, are all equal.¹¹ Once it was recognized that genes are (mostly) chromosomal segments, (as biologists discovered soon after the rediscovery of Mendel's laws), we understand that the law will not hold in general: alleles which are on the same chromosome (or, more exactly, close together on the same chromosome) will tend to be transmitted together because (ignoring recombination)¹² one member of each homologous pair is distributed to a gamete.¹³

Now it might seem that this is not very important. We could surely find a correct substitute for Mendel's second law by restricting the law so that it only talks about genes on nonhomologous chromosomes. Unfortunately, this will not quite do. There can be interference with normal cytological processes so that segregation of nonhomologous chromosomes need not be independent.¹⁴ However, my complaint about Mendel's second law is not that it is incorrect: many sciences use laws that are clearly recognized as approximations. Mendel's second law, amended or unamended, simply becomes irrelevant to subsequent research in classical genetics.

We envisaged amending Mendel's second law by using elementary principles of cytology, together with the identification of genes as chromosomal segments, to correct what was faulty in the unamended law. It is the fact that the application is so easy and that it can be carried out far more generally that makes the “law” it generates irrelevant. We can understand the transmission of genes by analyzing the cases that interest us

from a cytological perspective—by proceeding from “first principles,” as it were. Moreover, we can adopt this approach whether the organism is haploid, diploid or polyploid, whether it reproduces sexually or asexually, whether the genes with which we are concerned are or are not on homologous chromosomes, whether or not there is distortion of independent chromosomal segregation at meiosis. Cytology not only teaches us that the second law is false; it also tells us how to tackle the problem at which the second law was directed (the problem of determining frequencies for pairs of genes in gametes). The amended second law is a restricted statement of results obtainable using a general technique. What figures largely in genetics after Morgan is the technique, and this is hardly surprising when we realize that one of the major research problems of classical genetics has been the problem of discovering the distribution of genes *on the same chromosome*, a problem which is beyond the scope of the amended law.

Let us now turn from (R1) to (R2), assuming, contrary to what has just been argued, that we can identify the content of classical genetics with general principles about gene transmission. (Let us even suppose, for the sake of concreteness, that the principles in question are Mendel's laws—amended in whatever way the reductionist prefers.) To derive these principles from molecular biology, we need a bridge principle. I shall consider first statements of the form

$$(*) (x) (x \text{ is a gene} \leftrightarrow Mx)$$

where ‘ Mx ’ is an open sentence (possibly complex) in the language of molecular biology. Molecular biologists do not offer any appropriate statement. Nor do they seem interested in providing one. I claim that no appropriate bridge principle can be found.

Most genes are segments of DNA. (There are some organisms—viruses—whose genetic material is RNA; I shall henceforth ignore them.) Thanks to Watson and Crick, we know the molecular structure of DNA. Hence the problem of providing a statement of the above form becomes that of saying, in molecular terms, which segments of DNA count as genes.

Genes come in different sizes, and, for any given size, we can find segments of DNA of that size that are not genes. Therefore genes cannot be identified as segments of DNA containing a particular number of nucleotide pairs. Nor will it do to give a molecular characterization of those codons (triplets of nucleotides) that initiate and terminate transcription, and take a gene to be a segment of DNA between successive initiating and terminating codons. In the first place, mutation might produce a *single* allele containing within it codons for stopping and restarting transcription.¹⁵ Secondly, and much more importantly, the criterion is not general since not every gene is transcribed on mRNA.

The latter point is worth developing. Molecular geneticists recognize regulatory genes as well as structural genes. To cite a classic example, the

operator region in the *lac* operon of *E. coli* serves as a site for the attachment of protein molecules, thereby inhibiting transcription of mRNA and regulating enzyme production.¹⁶ Moreover, it is becoming increasingly obvious that genes are not always transcribed, but play a variety of roles in the economy of the cell.¹⁷

At this point, the reductionist may try to produce a bridge principle by brute force. Trivially, there are only a finite number of terrestrial organisms (past, present and future) and only a finite number of genes. Each gene is a segment of DNA with a particular structure and it would be possible, in principle, to provide a detailed molecular description of that structure. We can now give a molecular specification of the gene by enumerating the genes and disjoining the molecular descriptions.¹⁸ The point made above, that the segments which we count as genes do not share any structural property can now be put more precisely: any instantiation of (*) which replaces 'M' by a structural predicate from the language of molecular biology will insert a predicate that is essentially disjunctive.

Why does this matter? Let us imagine a reductionist using the enumerative strategy to deduce a general principle about gene transmission. After great labor, it is revealed that all actual genes satisfy the principle. I claim that more than this is needed to reduce a *law* about gene transmission. We envisage laws as sustaining counterfactuals, as applying to examples that might have been but which did not actually arise. To reduce the law it is necessary to show how possible but nonactual genes would have satisfied it. Nor can we achieve the reductionist's goal by adding further disjuncts to the envisaged bridge principle. For although there are only finitely many *actual* genes, there are indefinitely many genes which *might* have arisen.

At this point, the reductionist may protest that the deck has been stacked. There is no need to produce a bridge principle of the form (*). Recall that we are trying to derive a general law about the transmission of genes, whose paradigm is Mendel's second law. Now the gross logical form of Mendel's second law is:

$$(1) (x) (y) ((Gx \ \& \ Gy) \rightarrow Axy).$$

We might hope to obtain this from statements of the forms

$$(2) (x) (Gx \rightarrow Mx)$$

$$(3) (x) (y) ((Mx \ \& \ My) \rightarrow Axy)$$

where 'Mx' is an open sentence in the language of molecular biology. Now there will certainly be true statements of the form (2): for example, we can take 'Mx' as 'x is composed of DNA [or] x is composed of RNA'. The question is whether we can combine some such statement with other

appropriate premises—for example, some instance of (3)—so as to derive, and thereby explain (1). No geneticist or molecular biologist has advanced any suitable premises, and with good reason. We discover true statements of the form (2) by hunting for weak necessary conditions on genes, conditions which have to be met by genes but which are met by hordes of other biological entities as well. We can only hope to obtain *weak* necessary conditions because of the phenomenon that occupied us previously: from the molecular standpoint, genes are not distinguished by any common structure. Trouble will now arise when we try to show that the weak necessary condition is jointly sufficient for the satisfaction of the property (independent assortment at meiosis) that we ascribe to genes. The difficulty is illustrated by the example given above. If we take 'Mx' to be 'x is composed of DNA [or] x is composed of RNA' then the challenge will be to find a general law governing the distribution of all segments of DNA and RNA!

I conclude that (R2) is false. Reductionists cannot find the bridge principles they need, and the tactic of abandoning the form (*) for something weaker is of no avail. I shall now consider (R3). Let us concede both of the points that I have denied, allowing that there are general laws about the transmission of genes and that bridge principles are forthcoming. I claim that exhibiting derivations of the transmission laws from principles of molecular biology and bridge principles would not explain the laws, and, therefore, would not fulfill the major goal of reduction.

As an illustration, I shall use the envisaged amended version of Mendel's second law. Why do genes on nonhomologous chromosomes assort independently? Cytology provides the answer. At meiosis, chromosomes line up with their homologues. It is then possible for homologous chromosomes to exchange some genetic material, producing pairs of recombinant chromosomes. In the meiotic division, one member of each recombinant pair goes to each gamete, and the assignment of one member of one pair to a gamete is probabilistically independent of the assignment of a member of another pair to that gamete. Genes which occur close on the same chromosome are likely to be transmitted together (recombination is not likely to occur between them), but genes on nonhomologous chromosomes will assort independently.

This account is a perfectly satisfactory explanation of why our envisaged law is true to the extent that it is. (We recognize how the law could fail if there were some unusual mechanism linking particular nonhomologous chromosomes.) To emphasize the adequacy of the explanation is not to deny that it could be extended in certain ways. For example, we might want to know more about the mechanics of the process by which the chromosomes are passed on to the gametes. In fact, cytology provides such information. However, appeal to molecular biology would not deepen our understanding of the transmission law. Imagine a successful derivation of the law from principles of chemistry and a bridge principle

of the form (*). In charting the details of the molecular rearrangements the derivation would only blur the outline of a simple cytological story, adding a welter of irrelevant detail. Genes on nonhomologous chromosomes assort independently because nonhomologous chromosomes are transmitted independently at meiosis, and, so long as we recognize this, we do not need to know what the chromosomes are made of.

In explaining a scientific law, *L*, one often provides a deduction of *L* from other principles. Sometimes it is possible to explain some of the principles used in the deduction by deducing them, in turn, from further laws. Recognizing the possibility of a sequence of deductions tempts us to suppose that we could produce a better explanation of *L* by combining them, producing a more elaborate derivation in the language of our ultimate premises. But this is incorrect. What is relevant for the purposes of giving one explanation may be quite different from what is relevant for the purposes of explaining a law used in giving that original explanation. This general point is illustrated by the case at hand. We begin by asking why genes on nonhomologous chromosomes assort independently. The simple cytological story rehearsed above answers the question. That story generates *further* questions. For example, we might inquire why nonhomologous chromosomes are distributed independently at meiosis. To answer this question we would describe the formation of the spindle and the migration of chromosomes to the poles of the spindle just before meiotic division.¹⁹ Once again, the narrative would generate yet further questions. Why do the chromosomes "condense" at prophase? How is the spindle formed? Perhaps in answering these questions we would begin to introduce the chemical details of the process. Yet simply plugging a molecular account into the narratives offered at the previous stages would *decrease* the explanatory power of those narratives. What is relevant to answering our original question is the fact that nonhomologous chromosomes assort independently. What is relevant to the issue of why nonhomologous chromosomes assort independently is the fact that the chromosomes are not selectively oriented toward the poles of the spindle. (We need to eliminate the doubt that, for example, the paternal and maternal chromosomes become separated and aligned toward opposite poles of the spindle.) In neither case are the molecular details relevant. Indeed, adding those details would only disguise the relevant factor.

There is a natural reductionist response. The considerations of the last paragraphs presuppose far too subjective a view of scientific explanation. After all, even if we become lost in the molecular details, beings who are cognitively more powerful than we could surely recognize the explanatory force of the envisaged molecular derivation. However, this response misses a crucial point. The molecular derivation forfeits something important.

Recall the original cytological explanation. It accounted for the trans-

mission of genes by identifying meiosis as a process of a particular kind: a process in which paired entities (in this case, homologous chromosomes) are separated by a force so that one member of each pair is assigned to a descendant entity (in this case, a gamete). Let us call processes of this kind *PS-processes*. I claim first that explaining the transmission law requires identifying PS-processes as forming a natural kind to which processes of meiosis belong, and second that PS-processes cannot be identified as a kind from the molecular point of view.

If we adopt the familiar covering law account of explanation, then we shall view the cytological narrative as invoking a law to the effect that processes of meiosis are PS-processes and as applying elementary principles of probability to compute the distribution of genes to gametes from the laws that govern PS-processes. If the illumination provided by the narrative is to be preserved in a molecular derivation, then we shall have to be able to express the relevant laws as laws in the language of molecular biology, and this will require that we be able to characterize PS-processes as a natural kind from the molecular point of view. The same conclusion, to wit that the explanatory power of the cytological account can be preserved only if we can identify PS-processes as a natural kind in molecular terms, can be reached in analogous ways if we adopt quite different approaches to scientific explanation—for example, if we conceive of explanation as specifying causally relevant properties or as fitting phenomena into a unified account of nature.

However, PS-processes are heterogeneous from the molecular point of view. There are no constraints on the molecular structures of the entities which are paired or on the ways in which the fundamental forces combine to pair them and to separate them. The bonds can be forged and broken in innumerable ways: all that matters is that there be bonds that initially pair the entities in question and that are subsequently (somehow) broken. In some cases, bonds may be formed directly between constituent molecules of the entities in question; in others, hordes of accessory molecules may be involved. In some cases, the separation may occur because of the action of electromagnetic forces or even of nuclear forces; but it is easy to think of examples in which the separation is effected by the action of gravity. I claim, therefore, that PS-processes are realized in a motley of molecular ways. (I should note explicitly that this conclusion is independent of the issue of whether the reductionist can find bridge principles for the concepts of classical genetics.)

We thus obtain a reply to the reductionist charge that we reject the explanatory power of the molecular derivation simply because we anticipate that our brains will prove too feeble to cope with its complexities.²⁰ The molecular account objectively fails to explain because it cannot bring out that feature of the situation which is highlighted in the cytological story. It cannot show us that genes are transmitted in the ways that we

find them to be because meiosis is a PS-process and because any PS-process would give rise to analogous distributions. Thus (R3)—like (R1) and (R2)—is false.

3 | The Root of the Trouble

Where did we go wrong? Here is a natural suggestion. The most fundamental failure of reductionism is the falsity of (R1). Lacking an account of theories which could readily be applied to the cases of classical genetics and molecular genetics, the attempt to chart the relations between these theories was doomed from the start. If we are to do better, we must begin by asking a preliminary question: what is the structure of classical genetics?

I shall follow this natural suggestion, endeavoring to present a picture of the structure of classical genetics which can be used to understand the intertheoretic relations between classical and molecular genetics.²¹ As we have seen, the main difficulty in trying to axiomatize classical genetics is to decide what body of statements one is attempting to axiomatize. The history of genetics makes it clear that Morgan, Muller, Sturtevant, Beadle, McClintock, and others have made important contributions to genetic theory. But the statements occurring in the writings of these workers seem to be far too specific to serve as parts of a general theory. They concern the genes of particular kinds of organisms—primarily paradigm organisms, like fruit flies, bread molds, and maize. The idea that classical genetics is simply a heterogeneous set of statements about dominance, recessiveness, position effect, nondisjunction, and so forth, in *Drosophila*, *Zea mays*, *E. coli*, *Neurospora*, etc. flies in the face of our intuitions. The statements advanced by the great classical geneticists seem more like *illustrations* of the theory than *components* of it. (To know classical genetics it is not necessary to know the genetics of any particular organism, not even *Drosophila melanogaster*.) But the only alternative seems to be to suppose that there are general laws in genetics, never enunciated by geneticists but reconstructible by philosophers. At the very least, this supposition should induce the worry that the founders of the field, and those who write the textbooks of today, do a singularly bad job.

Our predicament provokes two main questions. First, if we focus on a particular time in the history of classical genetics, it appears that there will be a set of statements about inheritance in particular organisms, which constitutes the corpus which geneticists of that time accept: what is the relationship between this corpus and the version of classical genetic theory in force at the time? (In posing this question, I assume, contrary to fact, that the community of geneticists was always distinguished by unusual harmony of opinion; it is not hard to relax this simplifying assumption.) Second, we think of genetic theory as something that persisted through

various versions: what is the relation among the versions of classical genetic theory accepted at different times (the versions of 1910, 1930, and 1950, for example) which makes us want to count them as versions of the same theory?

We can answer these questions by amending a prevalent conception of the way in which we should characterize the state of a science at a time. The corpus of statements about the inheritance of characteristics accepted at a given time is only one component of a much more complicated entity that I shall call the *practice* of classical genetics at that time. There is a common language used to talk about hereditary phenomena, a set of accepted statements in that language (the corpus of beliefs about inheritance mentioned above), a set of questions taken to be the appropriate questions to ask about hereditary phenomena, and a set of patterns of reasoning which are instantiated in answering some of the accepted questions; (also: sets of experimental procedures and methodological rules, both designed for use in evaluating proposed answers; these may be ignored for present purposes). The practice of classical genetics at a time is completely specified by identifying each of the components just listed.²²

A pattern of reasoning is a sequence of *schematic sentences*, that is sentences in which certain items of nonlogical vocabulary have been replaced by dummy letters, together with a set of *filling instructions* which specify how substitutions are to be made in the schemata to produce reasoning which instantiates the pattern.²³ This notion of pattern is intended to explicate the idea of the common structure that underlies a group of problem-solutions.*

The foregoing definitions enable us to answer the two main questions I posed above. Beliefs about the particular genetic features of particular organisms illustrate or exemplify the version of genetic theory in force at the time in the sense that these beliefs figure in particular problem-solutions generated by the current practice. Certain patterns of reasoning are applied to give the answers to accepted questions, and, in making the application, one puts forward claims about inheritance in particular organisms. Classical genetics persists as a single theory with different versions at different times in the sense that different practices are linked by a chain of practices along which there are relatively small modifications in language, in accepted questions, and in the patterns for answering questions. In addition to this condition of historical connection, versions of classical genetic theory are bound by a common structure: each version uses certain

* For more on patterns of reasoning and their role in Kitcher's view of theoretical explanation as unification, see Philip Kitcher, "Explanatory Unification and the Causal Structure of the World," in P. Kitcher and W. C. Salmon, eds., *Scientific Explanation, Minnesota Studies in the Philosophy of Science*, vol. 13 (Minneapolis: University of Minnesota Press, 1989), 410–505, and Philip Kitcher, *The Advancement of Science* (Oxford: Oxford University Press, 1993), ch. 2.

expressions to characterize hereditary phenomena, accepts as important questions of a particular form, and offers a general style of reasoning for answering those questions. Specifically, throughout the career of classical genetics, the theory is directed toward answering questions about the distribution of characteristics in successive generations of a genealogy, and it proposes to answer those questions by using the probabilities of chromosome distribution to compute the probabilities of descendant genotypes.

The approach to classical genetics embodied in these answers is supported by reflection on what beginning students learn. Neophytes are not taught (and never have been taught) a few fundamental theoretical laws from which genetic “theorems” are to be deduced. They are introduced to some technical terminology, which is used to advance a large amount of information about special organisms. Certain questions about heredity in these organisms are posed and answered. Those who understand the theory are those who know what questions are to be asked about hitherto unstudied examples, who know how to apply the technical language to the organisms involved in these examples, and who can apply the patterns of reasoning which are to be instantiated in constructing answers. More simply, successful students grasp general patterns of reasoning which they can use to resolve new cases.

I shall now add some detail to my sketch of the structure of classical genetics, and thereby prepare the way for an investigation of the relations between classical genetics and molecular genetics. The initial family of problems in classical genetics, the family from which the field began, is the family of *pedigree problems*. Such problems arise when we confront several generations of organisms, related by specified connections of descent, with a given distribution of one or more characteristics. The question that arises may be to understand the given distribution of phenotypes, or to predict the distribution of phenotypes in the next generation, or to specify the probability that a particular phenotype will result from a particular mating. In general, classical genetic theory answers such questions by making hypotheses about the relevant genes, their phenotypic effects and their distribution among the individuals in the pedigree. Each version of classical genetic theory contains one or more problem-solving patterns exemplifying this general idea, but the detailed character of the pattern is refined in later versions, so that previously recalcitrant cases of the problem can be accommodated.

Each case of a pedigree problem can be characterized by a set of *data*, a set of *constraints*, and a question. In any example, the data are statements describing the distribution of phenotypes among the organisms in a particular pedigree, or a diagram conveying the same information. The level of detail in the data may vary widely: at one extreme we may be given a full description of the interrelationships among all individuals and the sexes of all those involved; or the data may only provide the numbers of individuals with specific phenotypes in each generation; or,

with minimal detail, we may simply be told that from crosses among individuals with specified phenotypes a certain range of phenotypes is found.

The constraints on the problem consist of general cytological information and descriptions of the chromosomal constitution of members of the species. The former will include the thesis that genes are (almost always)²⁴ chromosomal segments and the principles that govern meiosis. The latter may contain a variety of statements. It may be pertinent to know how the species under study reproduces, how sexual dimorphism is reflected at the chromosomal level, the chromosome number typical of the species, what loci are linked, what the recombination frequencies are, and so forth. As in the case of the data, the level of detail (and thus of stringency) in the constraints can vary widely.

Lastly, each problem contains a question that refers to the organisms described in the data. The question may take several forms: "What is the expected distribution of phenotypes from a cross between *a* and *b*?" (where *a*, *b* are specified individuals belonging to the pedigree described by the data), "What is the probability that a cross between *a* and *b* will produce an individual having *P*?" (where *a*, *b* are specified individuals of the pedigree described by the data and *P* is a phenotypic property manifested in this pedigree), "Why do we find the distribution of phenotypes described in the data?" and others.

Pedigree problems are solved by advancing pieces of reasoning that instantiate a small number of related patterns. In all cases the reasoning begins from a *genetic hypothesis*. The function of a genetic hypothesis is to specify the alleles that are relevant, their phenotypic expression, and their transmission through the pedigree. From that part of the genetic hypothesis that specifies the genotypes of the parents in any mating that occurs in the pedigree, together with the constraints on the problem, one computes the expected distribution of genotypes among the offspring. Finally, for any mating occurring in the pedigree, one shows that the expected distribution of genotypes among the offspring is consistent with the assignment of genotypes given by the genetic hypothesis.

The form of the reasoning can easily be recognized in examples—examples that are familiar to anyone who has ever looked at a textbook or a research report in genetics.²⁵ What interests me is the style of reasoning itself. The reasoning begins with a genetic hypothesis that offers four kinds of information: (a) Specification of the number of relevant loci and the number of alleles at each locus; (b) Specification of the relationships between genotypes and phenotypes; (c) Specification of the relations between genes and chromosomes, of facts about the transmission of chromosomes to gametes (for example, resolution of the question whether there is disruption of normal segregation) and about the details of zygote formation; (d) Assignment of genotypes to individuals in the pedigree. After showing that the genetic hypothesis is consistent with the data and constraints of the problem, the principles of cytology and the laws of prob-

ability are used to compute expected distributions of genotypes from crosses. The expected distributions are then compared with those assigned in part (d) of the genetic hypothesis.²⁶

Throughout the career of classical genetics, pedigree problems are addressed and solved by carrying out reasoning of the general type just indicated. Each version of classical genetic theory contains a pattern for solving pedigree problems with a method for computing expected genotypes which is adjusted to reflect the particular form of the genetic hypotheses that it sanctions. Thus one way to focus the differences among successive versions of classical genetic theory is to compare their conceptions of the possibilities for genetic hypotheses. As genetic theory develops, there is a changing set of conditions on admissible genetic hypotheses. Prior to the discovery of polygeny and pleiotropy (for example), part (a) of any adequate genetic hypothesis was viewed as governed by the requirement that there would be a one-one correspondence between loci and phenotypic traits.²⁷ After the discovery of incomplete dominance and epistasis, it was recognized that part (b) of an adequate hypothesis might take a form that had not previously been allowed: one is not compelled to assign to the heterozygote a phenotype assigned to one of the homozygotes, and one is also permitted to relativize the phenotypic effect of a gene to its genetic environment.²⁸ Similarly, the appreciation of phenomena of linkage, recombination, nondisjunction, segregation distortion, meiotic drive, unequal crossing over, and crossover suppression, modify conditions previously imposed on part (c) of any genetic hypothesis. In general, we can take each version of classical genetic theory to be associated with a set of conditions (usually not formulated explicitly) which govern admissible genetic hypotheses. While a general form of reasoning persists through the development of classical genetics, the patterns of reasoning used to resolve cases of the pedigree problem are constantly fine-tuned as geneticists modify their views about what forms of genetic hypothesis are allowable.

So far I have concentrated exclusively on classical genetic theory as a family of related patterns of reasoning for solving the pedigree problem. It is natural to ask if versions of the theory contain patterns of reasoning for addressing other questions. I believe that they do. The heart of the theory is the theory of *gene transmission*, the family of reasoning patterns directed at the pedigree problem. Out of this theory grow other subtheories. The theory of *gene mapping* offers a pattern of reasoning which addresses questions about the relative positions of loci on chromosomes. It is a direct result of Sturtevant's insight that one can systematically investigate the set of pedigree problems associated with a particular species. In turn, the theory of gene mapping raises the question of how to identify mutations, issues which are to be tackled by the *theory of mutation*. Thus we can think of classical genetics as having a central theory, the theory of gene transmission, which develops in the ways I have described above,

surrounded by a number of satellite theories that are directed at questions arising from the pursuit of the central theory. Some of these satellite theories (for example, the theory of gene mapping) develop in the same continuous fashion. Others, like the theory of mutation, are subject to rather dramatic shifts in approach.

4 | Molecular Genetics and Classical Genetics

Armed with some understanding of the structure and evolution of classical genetics, we can finally return to the question with which we began. What is the relation between classical genetics and molecular genetics? When we look at textbook presentations and the pioneering research articles that they cite, it is not hard to discern major ways in which molecular biology has advanced our understanding of hereditary phenomena. We can readily identify particular molecular explanations which illuminate issues that were treated incompletely, if at all, from the classical perspective. What proves puzzling is the connection of these explanations to the theory of classical genetics. I hope that the account of the last section will enable us to make the connection.

I shall consider three of the most celebrated achievements of molecular genetics. Consider first the question of *replication*. Classical geneticists believed that genes can replicate themselves. Even before the experimental demonstration that all genes are transmitted to all the somatic cells of a developing embryo, geneticists agreed that normal processes of mitosis and meiosis must involve gene replication. Muller's suggestion that the central problem of genetics is to understand how mutant alleles, incapable of performing wild-type functions in producing the phenotype, are nonetheless able to replicate themselves, embodies this consensus. Yet classical genetics had no account of gene replication. A molecular account was an almost immediate dividend of the Watson-Crick model of DNA.

Watson and Crick suggested that the two strands of the double helix unwind and each strand serves as the template for the formation of a complementary strand. Because of the specificity of the pairing of nucleotides, reconstruction of DNA can be unambiguously directed by a single strand. This suggestion has been confirmed and articulated by subsequent research in molecular biology.²⁹ The details are more intricate than Watson and Crick may originally have believed, but the outline of their story stands.

A second major illumination produced by molecular genetics concerns the characterization of mutation. When we understand the gene as a segment of DNA we recognize the ways in which mutant alleles can be produced. "Copying errors" during replication can cause nucleotides to

be added, deleted or substituted. These changes will often lead to alleles that code for different proteins, and which are readily recognizable as mutants through their production of deviant phenotypes. However, molecular biology makes it clear that there can be *hidden* mutations, mutations that arise through nucleotide substitutions that do not change the protein produced by a structural gene (the genetic code is redundant) or through substitutions that alter the form of the protein in trivial ways. The molecular perspective provides us with a general answer to the question, "What is a mutation?" namely that a mutation is the modification of a gene through insertion, deletion or substitution of nucleotides. This general answer yields a basic method for tackling (in principle) questions of form, "Is *a* a mutant allele?" namely a demonstration that *a* arose through nucleotide changes from alleles that persist in the present population. The method is frequently used in studies of the genetics of bacteria and bacteriophage, and can sometimes be employed even in inquiries about more complicated organisms. So, for example, there is good biochemical evidence for believing that some alleles which produce resistance to pesticides in various species of insects arose through nucleotide changes in the alleles naturally predominating in the population.³⁰

I have indicated two general ways in which molecular biology answers questions that were not adequately resolved by classical genetics. Equally obvious are a large number of more specific achievements. Identification of the molecular structures of particular genes in particular organisms has enabled us to understand why those genes combine to produce the phenotypes they do. One of the most celebrated cases is that of the normal allele for the synthesis of human hemoglobin and the mutant allele that is responsible for sickle-cell anemia.³¹ The hemoglobin molecule—whose structure is known in detail—is built up from four amino-acid chains (two "α-chains" and two "β-chains"). The mutant allele results from substitution of a single nucleotide with the result that one amino acid is different (the sixth amino acid in the β-chains). This slight modification causes a change in the interactions of hemoglobin molecules: deoxygenated mutant hemoglobin molecules combine to form long fibres. Cells containing the abnormal molecule become deformed after they have given up their oxygen, and because they become rigid, they can become stuck in narrow capillaries, if they give up their oxygen too soon. Individuals who are homozygous for the mutant gene are vulnerable to experience blockages of blood flow. However, in heterozygous individuals, there is enough normal hemoglobin in blood cells to delay the time of formation of the distorting fibres, so that the individual is physiologically normal.

This example is typical of a broad range of cases, among which are some of the most outstanding achievements of molecular genetics. In all of the cases, we replace a simple assertion about the existence of certain alleles which give rise to various phenotypes with a molecular character-

ization of those alleles from which we can derive descriptions of the phenotypes previously attributed.

I claim that the successes of molecular genetics which I have just briefly described—and which are among the accomplishments most emphasized in the biological literature—can be understood from the perspective on theories that I have developed above. The three examples reflect three different relations among successive theories, all of which are different from the classical notion of reduction (and the usual modifications of it). Let us consider them in turn.

The claim that genes can replicate does not have the status of a central law of classical genetic theory.³² It is not something that figures prominently in the explanations provided by the theory (as, for example, the Boyle-Charles law is a prominent premise in some of the explanations yielded by phenomenological thermodynamics). Rather, it is a claim that classical geneticists took for granted, a claim presupposed by explanations, rather than an explicit part of them. Prior to the development of molecular genetics that claim had come to seem increasingly problematic. If genes can replicate, how do they manage to do it? Molecular genetics answered the worrying question. It provided a theoretical demonstration of the possibility of an antecedently problematic presupposition of classical genetics.

We can say that a theory presupposes a statement *p* if there is some problem-solving pattern of the theory, such that every instantiation of the pattern contains statements that jointly imply the truth of *p*. Suppose that, at a given stage in the development of a theory, scientists recognize an argument from otherwise acceptable premises which concludes that it is impossible that *p*. Then the presupposition *p* is problematic for those scientists. What they would like would be an argument showing that it is possible that *p* and explaining what is wrong with the line of reasoning which appears to threaten the possibility of *p*. If a new theory generates an argument of this sort, then we can say that the new theory gives a theoretical demonstration of the possibility of an antecedently problematic presupposition of the old theory.

A less abstract account will help us to see what is going on in the case of gene replication. Very frequently, scientists take for granted in their explanations some general property of entities that they invoke. Their assumption can come to seem problematic if the entities in question are supposed to belong to a kind, and there arises a legitimate doubt about whether members of the kind can have the property attributed. A milder version of the problem arises if, in all cases in which the question of whether things of the general kind have the property can be settled by appealing to background theory, it turns out that the answer is negative. Under these circumstances, the scientists are committed to regarding their favored entities as unlike those things of the kind which are amenable to theoretical study with respect to the property under discussion. The situ-

ation is worse if background theory provides an argument for thinking that no things of the kind can have the property.

Consider now the case of gene replication. For any problem-solution offered by any version of the theory of gene transmission (the central subtheory of classical genetic theory), that problem-solution will contain sentences implying that the alleles which it discusses are able to replicate. Classical genetics presupposes that a large number of identifiable genes can replicate. This presupposition was always weakly problematic because genes were taken to be complicated molecules and, in all cases in which appeal to biochemistry could be made to settle the issue of whether a molecular structure was capable of replication, the issue was decided in the negative. Muller exacerbated the problem by suggesting that mutant alleles are damaged molecules (after all, many of them were produced through x-ray bombardment, an extreme form of molecular torture). So there appeared to be a strong argument against the possibility that any mutant allele can replicate. After the work of Watson, Crick, Kornberg, and others, there was a theoretical demonstration of the allegedly problematic possibility. One can show that genes can replicate by showing that any segment of DNA (or RNA) can replicate. (DNA and RNA are the genetic materials. Establishing the power of the genetic material to replicate bypasses the problem of deciding which segments are genes. Thus the difficulties posed by the falsity of [R2] are avoided.) The Watson-Crick model provides a characterization of the (principal) genetic material, and when this description is inserted into standard patterns of chemical reasoning one can generate an argument whose conclusion asserts that, under specified conditions, DNA replicates. Moreover, given the molecular characterization of DNA and of mutation, it is possible to see that although mutant alleles are "damaged" molecules, the kind of damage (insertion, deletion or substitution of nucleotides) does not affect the ability of the resultant molecule to replicate.

Because theoretical demonstrations of the possibility of antecedently problematic presuppositions involve derivation of conclusions of one theory from the premises supplied by a background theory, it is easy to assimilate them to the classical notion of reduction. However, on the account I have offered, there are two important differences. First, there is no commitment to the thesis that genetic theory can be formulated as (the deductive closure of) a conjunction of laws. Second, it is not assumed that all general statements about genes are equally in need of molecular derivation. Instead, one particular thesis, a thesis that underlies all the explanations provided by classical genetic theory, is seen as especially problematic, and the molecular derivation is viewed as addressing a specific problem that classical geneticists had already perceived. Where the reductionist identifies a general benefit in deriving all the axioms of the reduced theory, I focus on a particular derivation of a claim that has no title as an axiom of classical genetics, a derivation which responds to a particular

explanatory difficulty of which classical geneticists were acutely aware. The reductionist's global relation between theories does not obtain between classical and molecular genetics, but something akin to it does hold between special fragments of these theories.³³

The second principal achievement of molecular genetics, the account of mutation, involves a conceptual refinement of prior theory. Later theories can be said to provide conceptual refinements of earlier theories when the later theory yields a specification of entities that belong to the extensions of predicates in the language of the earlier theory, with the result that the ways in which the referents of these predicates are fixed are altered in accordance with the new specifications. Conceptual refinement may occur in a number of ways. A new theory may supply a descriptive characterization of the extension of a predicate for which no descriptive characterization was previously available; or it may offer a new description which makes it reasonable to amend characterizations that had previously been accepted.³⁴ In the case at hand, the referent of many tokens of 'mutant allele' was initially fixed through the description "chromosomal segment producing a heritable deviant phenotype." After Bridges's discovery of unequal crossing-over at the *Bar* locus in *Drosophila*, it was evident to classical geneticists that this descriptive specification covered cases in which the internal structure of a gene was altered and cases in which neighboring genes were transposed. Thus it was necessary to retreat to the less applicable description "chromosomal segment producing a heritable deviant phenotype as the result of an internal change within an allele." Molecular genetics offers a precise account of the internal changes, with the result that the description can be made more informative: mutant alleles are segments of DNA that result from prior alleles through deletion, insertion, or substitution of nucleotides. This re-fixing of the referent of 'mutant allele' makes it possible in principle to distinguish cases of mutation from cases of recombination, and thus to resolve those controversies that frequently arose from the use of 'mutant allele' in the later days of classical genetics.³⁵

Finally, let us consider the use of molecular genetics to illuminate the action of particular genes. Here we again seem to find a relationship that initially appears close to the reductionist's ideal. Statements that are invoked as premises in particular problem-solutions—statements that ascribe particular phenotypes to particular genotypes—are derived from molecular characterizations of the alleles involved. On the account of classical genetics offered in Section 3, each version of classical genetic theory includes in its schema for genetic hypotheses a clause which relates genotypes to phenotypes. . . . Generalizing from the hemoglobin example, we might hope to discover a pattern of reasoning within molecular genetics that would generate as its conclusion the schema for assigning phenotypes to genotypes.

It is not hard to characterize the relation just envisioned. Let us say

that a theory T' provides an *explanatory extension* of a theory T just in case there is some problem-solving pattern of T one of whose schematic premises can be generated as the conclusion of a problem-solving pattern of T' . When a new theory provides an explanatory extension of an old theory, then particular premises occurring in explanatory derivations given by the old theory can themselves be explained by using arguments furnished by the new theory. However, it does not follow that the explanations provided by the old theory can be improved by replacing the premises in question with the pertinent derivations. What is relevant for the purposes of explaining some statement S may not be relevant for the purposes of explaining a statement S' which figures in an explanatory derivation of S .

Even though reductionism fails, it may appear that we can capture part of the spirit of reductionism by deploying the notion of explanatory extension. The thesis that molecular genetics provides an explanatory extension of classical genetics embodies the idea of a global relationship between the two theories, while avoiding two of the three troubles that were found to beset reductionism. That thesis does not simply assert that some specific presupposition of classical genetics (for example, the claim that genes are able to replicate) can be derived as the conclusion of a molecular argument, but offers a general connection between premises of explanatory derivations in classical genetics and explanatory arguments from molecular genetics. It is formulated so as to accommodate the failure of (R1) and to honor the picture of classical genetics developed in Section 3. Moreover, the failure of (R2) does not affect it. If we take the hemoglobin example as a paradigm, we can justifiably contend that the explanatory extension does not require any general characterization of genes in molecular terms. All that is needed is the possibility of deriving phenotypic descriptions from molecular characterizations of the structures of *particular* genes. Thus, having surmounted two hurdles, our modified reductionist thesis is apparently within sight of success.

Nevertheless, even born-again reductionism is doomed to fall short of salvation. Although it is true that molecular genetics belongs to a cluster of theories which, taken together, provide an explanatory extension of classical genetics, molecular genetics, on its own, cannot deliver the goods. There are some cases in which the ancillary theories do not contribute to the explanation of a classical claim about gene action. In such cases, the classical claim can be derived and explained by instantiating a pattern drawn from molecular genetics. The example of human hemoglobin provides one such case. But this example is atypical.

Consider the way in which the hemoglobin example works. Specification of the molecular structures of the normal and mutant alleles, together with a description of the genetic code, enables us to derive the composition of normal and mutant hemoglobin. Application of chemistry then yields descriptions of the interactions of the proteins. With the aid

of some facts about human blood cells, one can then deduce that the sickling effect will occur in abnormal cells, and, given some facts about human physiology, it is possible to derive the descriptions of the phenotypes. There is a clear analogy here with some cases from physics. The assumptions about blood cells and physiological needs seem to play the same role as the boundary conditions about shapes, relative positions and velocities of planets that occur in Newtonian derivations of Kepler's laws. In the Newtonian explanation we can see the application of a general pattern of reasoning—the derivation of explicit equations of motion from specifications of the forces acting—which yields the general result that a body under the influence of a centrally directed inverse square force will travel in a conic section; the general result is then applied to the motions of the planets by incorporating pieces of astronomical information. Similarly, the derivation of the classical claims about the action of the normal and mutant hemoglobin genes can be seen as a purely chemical derivation of the generation of certain molecular structures and of the interactions among them. The chemical conclusions are then applied to the biological system under consideration by introducing three “boundary conditions”: first, the claim that the altered molecular structures only affect development to the extent of substituting a different molecule in the erythrocytes (the blood cells that transport hemoglobin); second, a description of the chemical conditions in the capillaries; and third, a description of the effects upon the organism of capillary blockage.

The example is able to lend comfort to reductionism precisely because of an atypical feature. In effect, one concentrates on the *differences* among the phenotypes, takes for granted the fact that in all cases development will proceed normally to the extent of manufacturing erythrocytes—which are, to all intents and purposes, simply sacks for containing hemoglobin molecules—and compares the difference in chemical effect of the cases in which the erythrocytes contain different molecules. *The details of the process of development can be ignored.* However, it is rare for the effect of a mutation to be so simple. Most structural genes code for molecules whose presence or absence make subtle differences. Thus, typically, a mutation will affect the distribution of chemicals in the cells of a developing embryo. A likely result is a change in the timing of intracellular reactions, a change that may, in turn, alter the shape of the cell. Because of the change of shape, the geometry of the embryonic cells may be modified. Cells that usually come into contact may fail to touch. Because of this, some cells may not receive the molecules necessary to switch on certain batteries of genes. Hence the chemical composition of these cells will be altered. And so it goes.³⁶

Quite evidently, in examples like this, (which include most of the cases in which molecular considerations can be introduced into embryology) the reasoning that leads us to a description of the phenotype associated with a genotype will be much more complicated than that found

in the hemoglobin case. It will not simply consist in a chemical derivation adapted with the help of a few boundary conditions furnished by biology. Instead, we shall encounter a sequence of subarguments: molecular descriptions lead to specifications of cellular properties, from these specifications we draw conclusions about cellular interactions, and from these conclusions we arrive at further molecular descriptions. There is clearly a pattern of reasoning here which involves molecular biology and which extends the explanations furnished by classical genetics by showing how phenotypes depend upon genotypes—but I think it would be folly to suggest that the extension is provided by molecular genetics alone.

In Section 2, we discovered that the traditional answer to the philosophical question of understanding the relation that holds between molecular genetics and classical genetics, the reductionist's answer, will not do. Section 3 attempted to build on the diagnosis of the ills of reductionism, offering an account of the structure and evolution of classical genetics that would improve on the picture offered by those who favor traditional approaches to the nature of scientific theories. In the present section, I have tried to use the framework of Section 3 to understand the relations between molecular genetics and classical genetics. Molecular genetics has done something important for classical genetics, and its achievements can be recognized by seeing them as instances of the intertheoretic relations that I have characterized. Thus I claim that the problem from which we began is solved.

So what? Do we have here simply a study of a particular case—a case which has, to be sure, proved puzzling for the usual accounts of scientific theories and scientific change? I hope not. Although the traditional approaches may have proved helpful in understanding some of the well-worn examples that have been the stock-in-trade of twentieth century philosophy of science, I believe that the notion of scientific practice sketched in Section 3 and the intertheoretic relations briefly characterized here will both prove helpful in analyzing the structure of science and the growth of scientific knowledge *even in those areas of science where traditional views have seemed most successful*.³⁷ Hence the tale of two sciences which I have been telling is not merely intended as a piece of local history that fills a small but troublesome gap in the orthodox chronicles. I hope that it introduces concepts of general significance in the project of understanding the growth of science.

5 | Anti-Reductionism and the Organization of Nature

One loose thread remains. The history of biology is marked by continuing opposition between reductionists and anti-reductionists. Reductionism

thrives on exploiting the charge that it provides the only alternative to the mushy incomprehensibility of vitalism. Anti-reductionists reply that their opponents have ignored the organismic complexity of nature. Given the picture painted above, where does this traditional dispute now stand?

I suggest that the account of genetics which I have offered will enable reductionists to provide a more exact account of what they claim, and will thereby enable anti-reductionists to be more specific about what they are denying. Reductionists and anti-reductionists agree in a certain minimal physicalism. To my knowledge, there are no major figures in contemporary biology who dispute the claim that each biological event, state or process is a complex physical event, state, or process. The most intricate part of ontogeny or phylogeny involves countless changes of physical state. What antireductionists emphasize is the organization of nature and the "interactions among phenomena at different levels." The appeal to organization takes two different forms. When the subject of controversy is the proper form of evolutionary theory, then anti-reductionists contend that it is impossible to regard all selection as operating at the level of the gene.³⁸ What concerns me here is not this area of conflict between reductionists and their adversaries, but the attempt to block claims for the hegemony of molecular studies in understanding the physiology, genetics, and development of organisms.³⁹

A sophisticated reductionist ought to allow that, in the current practice of biology, nature is divided into levels which form the proper provinces of areas of biological study: molecular biology, cytology, histology, physiology, and so forth. Each of these sciences can be thought of as using certain language to formulate the questions it deems important and as supplying patterns of reasoning for resolving those questions. Reductionists can now set forth one of two main claims. The stronger thesis is that the explanations provided by any biological theories can be reformulated in the language of molecular biology and be recast so as to instantiate the patterns of reasoning supplied by molecular biology. The weaker thesis is that molecular biology provides an explanatory extension of the other biological sciences.

Strong reductionism falls victim to the considerations that were advanced against (R3). The distribution of genes to gametes is to be explained, not by rehearsing the gory details of the reshuffling of the molecules, but through the observation that chromosomes are aligned in pairs just prior to the meiotic division, and that one chromosome from each matched pair is transmitted to each gamete. We may formulate this point in the biologists' preferred idiom by saying that the assortment of alleles is to be understood at the cytological level. What is meant by this description is that there is a pattern of reasoning which is applied to derive the description of the assortment of alleles and which involves predicates that characterize cells and their large-scale internal structures. That pattern of reasoning is to be objectively preferred to the molecular pattern which

would be instantiated by the derivation that charts that complicated rearrangements of individual molecules because it can be applied across a range of cases which would look heterogeneous from a molecular perspective. Intuitively, the cytological pattern makes connections which are lost at the molecular level, and it is thus to be preferred.

So far, anti-reductionism emerges as the thesis that there are *autonomous levels of biological explanation*. Anti-reductionism construes the current division of biology not simply as a temporary feature of our science stemming from our cognitive imperfections but as the reflection of levels of organization in nature. Explanatory patterns that deploy the concepts of cytology will endure in our science because we would forswear significant unification (or fail to employ the relevant laws, or fail to identify the causally relevant properties) by attempting to derive the conclusions to which they are applied using the vocabulary and reasoning patterns of molecular biology. But the autonomy thesis is only the beginning of anti-reductionism. A stronger doctrine can be generated by opposing the weaker version of sophisticated reductionism.

In Section 4, I raised the possibility that molecular genetics may be viewed as providing an explanatory extension of classical genetics through deriving the schematic sentence that assigns phenotypes to genotypes from a molecular pattern of reasoning. This apparent possibility fails in an instructive way. Anti-reductionists are not only able to contend that there are autonomous levels of biological explanation. They can also resist the weaker reductionist view that explanation always flows from the molecular level up. Even if reductionists retreat to the modest claim that, while there are autonomous levels of explanation, descriptions of cells and their constituents are always explained in terms of descriptions about genes, descriptions of tissue geometry are always explained in terms of descriptions of cells, and so forth, anti-reductionists can resist the picture of a unidirectional flow of explanation. Understanding the phenotypic manifestation of a gene, they will maintain, requires constant shifting back and forth across levels. Because developmental processes are complex and because changes in the timing of embryological events may produce a cascade of effects at several different levels, one sometimes uses descriptions at higher levels to explain what goes on at a more fundamental level.

For example, to understand the phenotype associated with a mutant limb-bud allele, one may begin by tracing the tissue geometry to an underlying molecular structure. The molecular constitution of the mutant allele gives rise to a nonfunctional protein, causing some abnormality in the internal structures of cells. The abnormality is reflected in peculiarities of cell shape, which, in turn, affects the spatial relations among the cells of the embryo. So far we have the unidirectional flow of explanation which the reductionist envisages. However, the subsequent course of the explanation is different. Because of the abnormal tissue geometry, cells that are normally in contact fail to touch; because they do not touch, certain im-

portant molecules, which activate some batteries of genes, do not reach crucial cells; because the genes in question are not “switched on” a needed morphogen is not produced; the result is an abnormal morphology in the limb.

Reductionists may point out, quite correctly, that there is some very complex molecular description of the entire situation. The tissue geometry is, after all, a configuration of molecules. But this point is no more relevant than the comparable claim about the process of meiotic division in which alleles are distributed to gametes. Certain genes are not expressed because of the geometrical structure of the cells in the tissue: *the pertinent cells are too far apart*. However this is realized at the molecular level, our explanation must bring out the salient fact that it is the presence of a gap between cells that are normally adjacent that explains the nonexpression of the genes. As in the example of allele transmission at meiosis, we lose sight of the important connections by attempting to treat the situation from a molecular point of view. As before, the point can be sharpened by considering situations in which radically different molecular configurations realize the crucial feature of the tissue geometry: situations in which heterogeneous molecular structures realize the breakdown of communication between the cells.

Hence, embryology provides support for the stronger anti-reductionist claim. Not only is there a case for the thesis of autonomous levels of explanation, but we find examples in which claims at a more fundamental level (specifically, claims about gene expression) are to be explained in terms of claims at a less fundamental level (specifically, descriptions of the relative positions of pertinent cells). Two anti-reductionist biologists put the point succinctly:

... a developmental program is not to be viewed as a linearly organized causal chain from genome to phenotype. Rather, morphology emerges as a consequence of an increasingly complex dialogue between cell populations, characterized by their geometric continuities, and the cells' genomes, characterized by their states of gene activity.⁴⁰

A corollary is that the explanations provided by the “less fundamental” biological sciences are not extended by molecular biology alone.

It would be premature to claim that I have shown how to reformulate the anti-reductionist appeals to the organization of nature in a completely precise way. My conclusion is that, to the extent that we can make sense of the present explanatory structure within biology—that division of the field into subfields corresponding to levels of organization in nature—we can also understand the anti-reductionist doctrine. In its minimal form, it is the claim that the commitment to several explanatory levels does not simply reflect our cognitive limitations; in its stronger form, it is the thesis that some explanations oppose the direction of preferred reductionistic

explanation. Reductionists should not dismiss these doctrines as incomprehensible mush unless they are prepared to reject as unintelligible the biological strategy of dividing the field (a strategy which seems to me well understood, even if unanalyzed).

The examples I have given seem to support both anti-reductionist doctrines. To clinch the case, further analysis is needed. The notion of explanatory levels obviously cries out for explication, and it would be illuminating to replace the informal argument that the unification of our beliefs is best achieved by preserving multiple explanatory levels with an argument based on a more exact criterion for unification. Nevertheless, I hope that I have said enough to make plausible the view that, despite the immense value of the molecular biology that Watson and Crick launched in 1953, molecular studies cannot cannibalize the rest of biology. Even if geneticists must become "physiological chemists" they should not give up being embryologists, physiologists, and cytologists.⁴¹

■ | Notes

1. "Variation due to change in the individual gene," reprinted in J. A. Peters ed., *Classic Papers in Genetics* (Englewood Cliffs, N.J.: Prentice-Hall, 1959), pp. 104–116. Citation from p. 115.
2. "Molecular Structure of Nucleic Acids," *Nature* 171 (1953), pp. 737–738; reprinted in Peters, op. cit., pp. 241–243. Watson and Crick amplified their suggestion in "Genetic Implications of the Structure of Deoxyribonucleic Acid" *Nature* 171 (1953), pp. 934–937.
3. The most sophisticated attempts to work out a defensible version of reductionism occur in articles by Kenneth Schaffner. See, in particular, "Approaches to Reduction," *Philosophy of Science* 34 (1967), pp. 137–147; "The Watson-Crick Model and Reductionism," *British Journal for the Philosophy of Science* 20 (1969), pp. 325–348; "The Peripherality of Reductionism in the Development of Molecular Biology," *Journal of the History of Biology* 7 (1974), pp. 111–139; and "Reductionism in Biology: Prospects and Problems," R. S. Cohen *et al.* eds., *PSA 1974*, (Boston: D. Reidel, 1976), pp. 613–632. See also Michael Ruse, "Reduction, Replacement, and Molecular Biology," *Dialectica* 25 (1971), pp. 38–72; and William K. Goossens, "Reduction by Molecular Genetics," *Philosophy of Science* 45 (1978), pp. 78–95. A variety of anti-reductionist points are made in David Hull, "Reduction in Genetics—Biology or Philosophy?" *Philosophy of Science* 39 (1972), pp. 491–499; and Chapter 1 of *Philosophy of Biological Science* (Englewood Cliffs, N.J.: Prentice-Hall, 1974); in Steven Orla Kimbrough, "On the Reduction of Genetics to Molecular Biology," *Philosophy of Science* 46 (1979), pp. 389–406 and in Ernst Mayr, *The Growth of Biological Thought* (Cambridge, Mass.: Harvard University Press, 1982), pp. 59–63.
4. Typically, though not invariably. In a suggestive essay, "Reductive Explanation: A Functional Account," (R. S. Cohen *et al.* op. cit. pp. 671–710), William Wim-

satt offers a number of interesting ideas about intertheoretic relations and the case of genetics. Also provocative are Nancy Maull's "Unifying Science Without Reduction," *Studies in the History and Philosophy of Science* 8 (1977), pp. 143–171; and Lindley Darden and Nancy Maull, "Interfield Theories," *Philosophy of Science* 44 (1977), pp. 43–64. My chief complaint about the works I have cited is that unexplained technical notions—"mechanisms," "levels," "domain," "field," "theory"—are invoked (sometimes in apparently inconsistent ways), so that no precise answer to the philosophical problem posed in the text is ever given. Nevertheless, I hope that the discussion of the later sections of this paper will help to articulate more fully some of the genuine insights of these authors, especially those contained in Wimsatt's rich essay.

5. E. Nagel, *The Structure of Science* (New York: Harcourt Brace, 1961), Chapter 11. A simplified presentation can be found in Chapter 8 of C. G. Hempel, *Philosophy of Natural Science* (Englewood Cliffs, N.J.: Prentice-Hall, 1966).

6. Quite evidently, this is a weak version of what was once the "received view" of scientific theories, articulated in the works of Nagel and Hempel cited in the previous note. A sustained presentation and critique of the view is given in the Introduction to F. Suppe, ed., *The Structure of Scientific Theories* (Urbana: University of Illinois Press, 1973). The fact that the standard model of reduction presupposes the thesis that theories are reasonably regarded as sets of statements has been noted by Clark Glymour, "On Some Patterns of Reduction," *Philosophy of Science* 36 (1969), pp. 340–353, 342; and by Jerry Fodor (*The Language of Thought*, New York: Crowell, 1975, p. 11, footnote 10). Glymour endorses the thesis; Fodor is skeptical about it.

7. Philosophers often suggest that, in reduction, one derives *corrected* laws of the reduced theory from an *unmodified* reducing theory. But this is not the way things go in the paradigm cases: one doesn't correct Galileo's law by using Newtonian mechanics; instead, one neglects "insignificant terms" in the Newtonian equation of motion for a body falling under the influence of gravity; similarly, in deriving the Boyle-Charles law from kinetic theory (or statistical mechanics), it is standard to make idealizing assumptions about molecules, and so obtain the exact version of the Boyle-Charles law; subsequently, corrected versions are generated by "subtracting" the idealizing procedures. Although he usually views reduction as deriving a corrected version of the reduced theory, Schaffner notes that reduction might sometimes proceed by modifying the reducing theory ("Approaches to Reduction," p. 138; "The Watson-Crick Model and Reductionism" p. 322). In fact, the point was already made by Nagel, op. cit.

8. In part, because modification might produce an inconsistent theory that would permit the derivation of anything. In part, because of the traditionally vexing problem of the proper form for bridge principles in heterogeneous reductions in physics. The former problem is discussed in Glymour, op. cit., p. 352 and in Dudley Shapere, "Notes towards a Post-Positivistic Interpretation of Science" (P. Achinstein and S. Barker, eds. *The Legacy of Logical Positivism*, Baltimore: Johns Hopkins University Press, 1971). For discussion of the latter issue, see Larry Sklar, "Types of Inter-Theoretic Reduction," *British Journal for the Philosophy of Science* 18 (1967), pp. 109–124; Robert Causey, "Attribute Identities in Microreductions,"

Journal of Philosophy 69 (1972), pp. 407–422; and Berent Enç, “Identity Statements and Micro-Reductions,” *Journal of Philosophy* 73 (1976), pp. 285–306. The concerns that I shall raise are orthogonal to these familiar areas of dispute.

9. Hull, *Philosophy of Biological Science*, p. 23, adapted from Theodosius Dobzhansky, *Genetics of the Evolutionary Process* (New York: Columbia University Press, 1970), p. 167. Similarly molecular genetics is said to have the task of “discovering how molecularly characterized genes produce proteins which in turn combine to form gross phenotypic traits” (Hull *ibid.*: see also James D. Watson, *Molecular Biology of the Gene*, Menlo Park, Ca., W. A. Benjamin, 1976, p. 54).

10. The phenotype/genotype distinction was introduced to differentiate the observable characteristics of an organism from the underlying genetic factors. In subsequent discussions the notion of phenotype has been extended to include properties which are not readily observable (for example, the capacity of an organism to metabolize a particular amino acid). The expansion of the concept of phenotype is discussed in my paper “Genes,” *British Journal for the Philosophy of Science* 33 (1982), pp. 337–359.

11. A *locus* is the place on a chromosome occupied by a gene. Different genes which can occur at the same locus are said to be *alleles*. In diploid organisms, chromosomes line up in pairs just before the meiotic division that gives rise to gametes. The matched pairs are pairs of *homologous chromosomes*. If different alleles occur at corresponding loci on a pair of homologous chromosomes, the organism is said to be *heterozygous* at these loci.

12. *Recombination* is the process (which occurs before meiotic division) in which a chromosome exchanges material with the chromosome homologous with it. Alleles which occur on one chromosome may thus be transferred to the other chromosome, so that new genetic combinations can arise.

13. Other central Mendelian claims also turn out to be false. The Mendelian principle that if an organism is heterozygous at a locus then the probabilities of either allele being transmitted to a gamete are equal falls afoul of cases of meiotic drive. (A notorious example is the *t*-allele in the house mouse, which is transmitted to 95% of the sperm of males who are heterozygous for it and the wild-type allele; see R. C. Lewontin and L. C. Dunn, “The Evolutionary Dynamics of a Polymorphism in the House Mouse,” *Genetics* 45 (1960), pp. 705–722.) Even the idea that genes are transmitted across the generations, unaffected by their presence in intermediate organisms, must be given up once we recognize that intra-allelic recombination can occur.

14. To the best of my knowledge, the mechanisms of this interference are not well understood. For a brief discussion, see J. Sybenga, *General Cytogenetics* (North-Holland, 1972), pp. 313–314. In this paper, I shall use “segregation distortion” to refer to cases in which there is a propensity for nonhomologous chromosomes to assort together. “Meiotic drive” will refer to examples in which one member of a pair of homologous chromosomes has a greater probability of being transmitted to a gamete. The literature in genetics exhibits some variation in the use of these terms. Let me note explicitly that, on these construals, both segregation distortion and meiotic drive will be different from *nondisjunction*, the

process in which a chromosome together with the whole (or a part) of the homologous chromosome is transmitted to a gamete.

15. This point raises some interesting issues. It is common practice in genetics to count a segment of DNA as a single gene if it was produced by mutation from a gene. Thus many mutant alleles are viewed as DNA segments in which modification of the sequence of bases has halted transcription too soon, with the result that the gene product is truncated and nonfunctional. My envisaged case simply assumes that a second mutation occurs further down the segment so that transcription starts and stops in two places, generating two useless gene products. The historical connection with the original allele serves to identify the segment as one gene.

Conversely, where there is no historical connection to any organism, one may have qualms about counting a DNA segment as a gene. Suppose that, in some region of space, a quirk of nature brings together the constituent atoms for the white eye mutant in *Drosophila melanogaster*, and that the atoms become arranged in the right way. Do we have here a *Drosophila* gene? If the right answer is "No" then it would seem that a molecular structure only counts as a gene given an appropriate history. I hasten to add that "appropriate histories" need not simply involve the usual biological ways in which organisms transmit, replicate and modify genes: one can reasonably hope to synthesize genes in the laboratory. The case seems analogous to questions that arise about personal identity. If a person's psychological features are replicated by a process that sets up the "right sort of causal connection" between person and product, then we are tempted to count the product as the surviving person. Similarly, if a molecular structure is generated in a way that sets up "the right sort of causal connection" between the structure and some prior gene then it counts as a gene. In both cases, causal connections of "the right sort" may be set up in everyday biological ways and by means of deliberate attempts to replicate a prior structure.

16. So called *structural genes* direct the formation of proteins by coding for RNA molecules. They are "transcribed" to produce *messenger* RNA (mRNA) which serves as a more immediate "blueprint" for the construction of the protein. Transcription is started and stopped through the action of regulatory genes. In the simplest regulatory system (that of the *lac* operon) an area adjacent to the structural gene serves as a "dumping ground" for a molecule. When concentration of the protein product becomes too high, the molecule attaches to this site and transcription halts; when more protein is required, the cell produces a molecule that removes the inhibiting molecule from the neighborhood of the structural gene, and transcription begins again. (For much more detail, see Watson, *op. cit.*, Chapter 14, and M. W. Strickberger, *Genetics* (New York: Macmillan, 1976), Chapter 29.)

17. The situation is complicated by the existence of "introns"—segments within genes whose products under transcription are later excised—and by the enormous amount of repetitive DNA that most organisms seem to contain. Moreover, the regulatory systems in eukaryotes appear to be much more complicated than the prokaryote systems (of which the *lac* operon is one paradigm). For a review of the situation, as of a few years ago, see Eric H. Davidson, *Gene Expression in Early Development* (New York: Academic Press, 1976).

18. The account will be even more complicated if we honor the suggestion of footnote 15, and suppose that, for a molecular structure to count as a gene, it must be produced in the right way.

19. Early in the process preceding meiotic division the chromosomes become more compact. As meiosis proceeds, the nucleus comes to contain a system of threads that resembles a spindle. Homologous chromosomes line up together near the center of the spindle, and they are oriented so that one member of each pair is slightly closer to one pole of the spindle, while the other is slightly closer to the opposite pole.

20. The point I have been making is related to an observation of Hilary Putnam's. Discussing a similar example, Putnam writes: "The same explanation will go in any world (whatever the microstructure) in which those *higher level structural features* are present"; he goes on to claim that "explanation is superior not just subjectively but methodologically. . . . if it brings out relevant laws." (Putnam, "Philosophy and our Mental Life," in *Mind, Language, and Reality*, (Cambridge, Cambridge University Press, 1975), pp. 291–303, p. 296). The point is articulated by Alan Garfinkel (*Forms of Explanation*, New Haven: Yale University Press, 1981), and William Wimsatt has also raised analogous considerations about explanation in genetics.

It is tempting to think that the independence of the "higher level structural features" in Putnam's example and in my own can be easily established: one need only note that there are worlds in which the same feature is present without any molecular realization. So, in the case discussed in the text, PS-processes might go on in worlds where all objects were perfect continua. But although this shows that PS-processes form a kind which could be realized without molecular reshufflings, we know that all *actual* PS-processes do involve such reshufflings. The reductionist can plausibly argue that *if* the set of PS-processes with molecular realizations is itself a natural kind, then the explanatory power of the cytological account can be preserved by identifying meiosis as a process of this narrower kind. Thus the crucial issue is not whether PS-processes form a kind with nonmolecular realizations, but whether those PS-processes which have molecular realizations form a kind that can be characterized from the molecular point of view. Hence, the easy strategy of responding to the reductionist must give way to the approach adopted in the text. (I am grateful to the editors of *The Philosophical Review* for helping me to see this point.)

21. It would be impossible in the scope of this paper to do justice to the various conceptions of scientific theory that have emerged from the demise of the "received view." Detailed comparison of the perspective I favor with more traditional approaches (both those that remain faithful to core ideas of the "received view" and those that adopt the "semantic view" of theories) must await another occasion.

22. My notion of a practice owes much to some neglected ideas of Sylvain Bromberger and Thomas Kuhn. See, in particular, Bromberger, "A Theory about the Theory of Theory and about the Theory of Theories," (W. L. Reese ed., *Philosophy of Science, The Delaware Seminar*, New York, 1963); and "Questions," (*Journal of Philosophy* 63 (1966), pp. 597–606); and Kuhn, *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1962) Chapters II–V. The relation

between the notion of a practice and Kuhn's conception of a paradigm is discussed in Chapter 7 of my book *The Nature of Mathematical Knowledge* (New York: Oxford University Press, 1983).

23. More exactly, a general argument pattern is a triple consisting of a sequence of schematic sentences (a *schematic argument*), a set of filling instructions (directions as to how dummy letters are to be replaced), and a set of sentences describing the inferential characteristics of the schematic argument (a *classification* for the schematic argument). A sequence of sentences instantiates the general argument pattern just in case it meets the following conditions: (i) the sequence has the same number of members as the schematic argument of the general argument pattern; (ii) each sentence in the sequence is obtained from the corresponding schematic sentence in accordance with the appropriate filling instructions; (iii) it is possible to construct a chain of reasoning which assigns to each sentence the status accorded to the corresponding schematic sentence by the classification. For some efforts at explanation and motivation, see my "Explanatory Unification," *Philosophy of Science* 48 (1981), pp. 507–531.

24. Sometimes particles in the cytoplasm account for hereditary traits. See Strickberger, *op. cit.*, pp. 257–265.

25. For examples, see Strickberger *op. cit.* Chapters 6–12, 14–17, especially Chapter 11; Peters, *op. cit.*; and H. L. K. Whitehouse, *Towards an Understanding of the Mechanism of Heredity* (London: Arnold, 1965).

26. The comparison will make use of standard statistical techniques, such as the chi-squared test.

27. *Polygeny* occurs when many genes affect one characteristic; *pleiotropy* occurs when one gene affects more than one characteristic.

28. *Incomplete dominance* occurs when the phenotype of the heterozygote is intermediate between that of the homozygotes; *epistasis* occurs when the effect of a particular combination of alleles at one locus depends on what alleles are present at another locus.

29. See Watson, *op. cit.*, Chapter 9; and Arthur Kornberg *DNA Synthesis* (San Francisco: W. H. Freeman, 1974).

30. See G. P. Georgioui, "The Evolution of Resistance to Pesticides," *Annual Review of Ecology and Systematics* 3 (1972), pp. 133–168.

31. See Watson, *op. cit.*, pp. 189–193; and T. H. Maugh II, "A New Understanding of Sickle Cell Emerges," *Science* 211 (1981), pp. 265–267.

32. However, one might claim that "Genes can replicate" is a law of genetics, in that it is general, lawlike, and true. This does not vitiate my claim that the structure of classical genetics is not to be sought by looking for a set of general laws, for the law in question is so weak that there is little prospect of finding supplementary principles which can be conjoined with it to yield a representation of genetic theory. I suggest that "Genes can replicate" is analogous to the thermodynamic "law," "Gases can expand," or to the Newtonian "law," "Forces can be combined." If the only laws that we could find in thermodynamics and mechanics were weak

statements of this kind we would hardly be tempted to conceive of these sciences as sets of laws. I think that the same point goes for genetics.

33. A similar point is made by Kenneth Schaffner in [*Discovery and Explanation in Biology and Medicine* (Chicago: University of Chicago Press, 1993)]. Schaffner's terminology is different from my own, and he continues to be interested in the prospects of global reduction, but there is considerable convergence between the conclusions that he reaches and those that I argue for in the present section.

34. There are numerous examples of such modifications from the history of chemistry. I try to do justice to this type of case in "Theories, Theorists, and Theoretical Change," *The Philosophical Review* 87 (1978), pp. 519–547 and in "Genes."

35. Molecular biology also provided significant refinement of the terms 'gene' and 'allele'. See "Genes."

36. For examples, see N. K. Wessels, *Tissue Interactions and Development* (Menlo Park, Ca.: W. A. Benjamin, 1977), especially Chapters 6, 7, 13–15; and Donald Ede, *An Introduction to Developmental Biology* (London: Blackie, 1978) especially Chapter 13.

37. I attempt to show how the same perspective can be fruitfully applied to other examples in "Explanatory Unification," Sections 3 and 4; *Abusing Science* (Cambridge: MIT Press, 1982) Chapter 2; and "Darwin's Achievement," [in Nicholas Rescher, ed., *Reason and Rationality in Science* (Washington, D.C.: University Press of America, 1985) 123–85.]

38. The extreme version of reductionism is defended by Richard Dawkins in *The Selfish Gene* (New York: Oxford University Press, 1976) and *The Extended Phenotype* (San Francisco: W. H. Freeman, 1982). For an excellent critique, see Elliott Sober and Richard C. Lewontin, "Artifact, Cause, and Genic Selection," *Philosophy of Science* 49 (1982), pp. 157–180. More ambitious forms of anti-reductionism with respect to evolutionary theory are advanced in S. J. Gould, "Is a new and general theory of evolution emerging?" *Paleobiology*, 6 (1980), pp. 119–130; N. Eldredge and J. Cracraft, *Phylogenetic Patterns and the Evolutionary Process* (New York: Columbia University Press, 1980); and Steven M. Stanley, *Macroevolution* (San Francisco: W. H. Freeman, 1979). A classic early source of some (but not all) later anti-reductionist themes is Ernst Mayr's *Animal Species and Evolution* (Cambridge, Harvard University Press, 1963) especially Chapter 10.

39. Gould's *Ontogeny and Phylogeny* (Harvard, 1977) provides historical illumination of both areas of debate about reductionism. Contemporary anti-reductionist arguments about embryology are expressed by Wessels (op. cit.) and Ede (op. cit.). See also G. Oster and P. Alberch, "Evolution and Bifurcation of Developmental Programs," *Evolution* 36 (1982), pp. 444–459.

40. Oster and Alberch, op. cit., p. 454. The diagram on p. 452 provides an equally straightforward account of their anti-reductionist position.

41. Earlier versions of this paper were read at Johns Hopkins University and at the University of Minnesota, and I am very grateful to a number of people for comments and suggestions. In particular, I would like to thank Peter Achinstein, John Beatty, Barbara Horan, Patricia Kitcher, Richard Lewontin, Kenneth Schaffner, William Wimsatt, an anonymous reader and the editors of *The Philosophical*

Review, all of whom have had an important influence on the final version. Needless to say, these people should not be held responsible for residual errors. I am also grateful to the American Council of Learned Societies and the Museum of Comparative Zoology at Harvard University for support and hospitality while I was engaged in research on the topics of this paper.

- 8.1 Nagel on the Logic of Reduction** 1005
Homogeneous and Inhomogeneous Reduction 1007
Nagel's Correspondence Analysis of Inhomogeneous Reduction 1009
Bridge Laws 1009
- 8.2 Feyerabend's Criticisms of Nagel's Model of Reduction** 1011
The Apparent Unreasonableness of Consistency and Meaning Invariance 1012
The Apparent Violation of the Consistency Condition 1016
The Apparent Violation of the Condition of Meaning Invariance 1017
- 8.3 Nickles on Two Concepts of Intertheoretic Reduction** 1022
Reevaluating Nagel's Account 1024
- 8.4 Kitcher on Reduction, Classical Genetics, and Molecular Biology** 1028
The Apparent Failure of Reductionism in Genetics 1030
Kitcher's Diagnosis of the Failure of Reductionism 1034
Presuppositions: The Example of Gene Replication 1036
Conceptual Refinement: The Example of Mutation 1037
Explanatory Extension: The Example of Sickle-Cell Anemia 1037
- 8.5 Summary** 1039

8 | COMMENTARY

8.1 | Nagel on the Logic of Reduction

As Ernest Nagel points out in the opening paragraphs of “Issues in the Logic of Reductive Explanations,” science has been and continues to be marked by strong reductive tendencies. Many of the classic cases of scientific reduction have brought about remarkable advances in our understanding of the world. For example, the reduction of classical thermodynamics to statistical mechanics brought with it new insights into the nature of heat and entropy; likewise, the reduction of optics to electromagnetic theory deepened our understanding of light and the behavior of light rays. The reductive explanation of chemical bonding by quantum mechanics revolutionized organic chemistry and biochemistry by revealing the nature of the covalent bond and the forces that determine the shape of molecules. Some people think that genetics and psychology are in the process of being reduced to biochemistry, and they expect great things as a consequence. Nagel wants to understand these and other classic cases of scientific reduction by articulating the structure of what he calls *reductive explanations*.

Before exploring Nagel’s analysis of reduction, two preliminary points are worth highlighting. First, however much the contrast between commonsense and scientific descriptions of the natural world might recommend it, we should avoid approaching the issue of reduction in terms of a distinction between appearance and reality. Although used by philosophers in the past, the appearance-reality distinction is too imprecise to bear the weight of an analysis of reduction. And in any case, the distinction is not of the sort we need. For the relation of chemistry to quantum mechanics, or thermodynamics to kinetic theory and statistical mechanics, is not one of appearance to reality. Even if some of our prescientific views of the natural world turn out to be false, even if they prove to be less about reality and more about mere appearance, nevertheless many of our commonsense judgments are true. Thus “steam is hotter than ice” does not report the full facts about steam and ice by virtue of which that sentence is true, but for all that, it can scarcely be reckoned a judgment of mere appearance.

Second, whatever its details, reduction properly so-called is typically construed as a relation (or family of relations) between some reducing theory T' and a reduced theory T . While it would take us too far afield here to discuss their exact nature, theories are commonly regarded as a set of statements—axioms, laws, empirical hypotheses—about the world. However natural it may be, then, to speak of certain phenomena, prop-

erties, or objects as reducing to the various phenomena, properties, or objects described by a reducing theory, it is strictly speaking the theories themselves, or particular expressions or concepts figuring in them, that stand in the reduction relation. Thus, when we find ourselves saying, for example, that heat reduces to the random motion of molecules, this is best regarded as an elliptical, shorthand expression of the view that *claims* about heat reduce to *claims* about molecular motion.

The second point just noted leads us to the core elements of reduction. Nagel has these elements in mind when he refers to the logic (or elsewhere, the formal conditions) of reduction.¹ As Nagel's talk of *reductive explanation* suggests, reducibility is commonly viewed as a kind of explanatory relation, whereby the reduced, secondary science may be accounted for as a special case or branch of some more general or inclusive primary science. Thus, unless the fundamental axioms, laws, and hypotheses of a science have been explicitly formulated as statements and their constituent terms given determinate meanings, it will be impossible to say whether some scientific theory does in fact reduce to another. This is particularly so given the deductive-nomological model of explanation that Nagel, Hempel, and many others accept. For according to that model, explanations are arguments (see chapter 6 above). Accordingly, the reduced and reducing theories must be expressed in meaningful statements from which the premises and conclusion of an explanatory reduction may be drawn. On Nagel's account, a reduction is made when the fundamental claims of a secondary science or theory *T* are shown to be logical consequences of the fundamental claims of the primary, reducing theory *T'*. The conclusions of such arguments are said to be reduced statements, while the premises are called reducing statements. This, then, is the core idea in Nagel's treatment of intertheoretic reduction—that the reduced claims of the secondary theory *T* be logically derivable from premises made up wholly of claims from the reducing, primary theory *T'*.

The central feature of Nagel's account of reduction—logical derivability—has two important consequences. First, the claims of *T* and *T'* must be logically consistent: since nothing incompatible with the fundamental claims of a theory can be logically derived from them, the claims of a reduced theory *T* must be consistent with those of the reducing theory *T'*. Second, any term shared by the reduced and reducing theory must have a common meaning if it appears in both the premises and the conclusion of a reductive explanation. For example, suppose that two sets of laws, one set from thermodynamics and the other from mechanics, both contain the terms *volume* and *pressure*. We could reduce thermodynamics to molecular mechanics using those laws only if *volume* and *pressure* have the same sense in both theories. If they do not, the supposed derivation would be invalid.²

HOMOGENEOUS AND INHOMOGENEOUS REDUCTION

The requirement of logical derivability raises an interesting question. Typically, reduction is associated with scientific progress or change—with cases in which an existing theory established in one domain of inquiry can be explained by a theory developed for some other, broader domain. The subject matter of the secondary, reduced theory would, in such cases, appear to be qualitatively distinct from that of the primary, reducing theory. Indeed, many of the concepts figuring in classical thermodynamics (e.g., heat and temperature) make no appearance as such in statistical mechanics; much of the descriptive vocabulary in chemistry is absent from quantum mechanics; and so on. But given that a secondary, reduced theory T employs in its fundamental laws predicates unique to that science, predicates that appear nowhere in the basic terminology of the primary, reducing theory T' , it is unclear how we can understand reduction as an explanatory derivation of the former theory from the latter. Irrelevant exceptions aside, no term can appear in the conclusion of a valid argument unless it is contained somewhere in the premises.³

Following Nagel, it has become common to distinguish “easy” cases of reduction from the “hard” cases. In the easy cases, the descriptive or subject-matter terms of the reduced theory are either present in or can be explicitly defined by terms in the reducing theory. In the hard cases, some term or terms in the reduced theory are neither present in nor definable by other terms in the reducing theory. The former cases are called *homogeneous reductions*, the latter—*inhomogeneous reductions* (or, as some authors prefer, *heterogeneous reductions*). A brief review of Nagel's account of homogeneous reduction will suffice, before we turn our attention to the more difficult and interesting case of inhomogeneous reduction.

A reduction is homogeneous if the descriptive vocabulary of the reduced theory T is a proper subset of the vocabulary of the reducing theory T' . Now all else being equal, were T simply a fragment of T' , straightforwardly deducible from T' as a mere part of it, it would be implausible to regard the relationship between the two as reductive. Typical examples of homogeneous reduction are, rather, cases in which an earlier theory T , developed as an independent account of some domain of natural phenomena, comes to be explained as a consequence of a later, newer theory T' of some broader domain of phenomena. Thus, Galileo's law of falling bodies and Kepler's laws of planetary motion may be said to reduce to Newtonian mechanics and gravitational theory: not only do the fundamental concepts of those earlier-formulated laws (distance, velocity, acceleration, etc.) appear in Newton's own account, but the restricted and apparently dissimilar kinds of motion to which Galileo's and Kepler's laws apply (terrestrial, celestial) are also given a fundamentally unified treatment by Newton's general account of motion. Therefore, if the laws of

terrestrial free fall and celestial motion can be derived from Newton's theory, then Nagel's formal conditions for a homogeneous reduction are met and those laws are said to be reducible to Newton's theory. (We shall take up some difficulties with this account presently.)

Inhomogeneous reduction occurs when the descriptive vocabulary of the reduced theory T is not a proper subset of the vocabulary of the reducing theory T' . We already encountered a standard example of inhomogeneous reduction—the explanatory reduction of thermodynamics, with its distinctive concepts of heat and temperature, to statistical mechanics. And we noted, too, the main respect in which an analysis of such a reduction is less straightforward than that of homogeneous cases: inhomogeneous reductions look unfit for the deductive explanation model, since it is unclear how the reduced laws in the conclusion of such an explanatory argument could contain terms or concepts absent from the reducing premises. If reduction is to be construed on the deductive model, and if inhomogeneous cases are a species of reduction, then it would seem that the premises of such reductive arguments must be supplemented with assumptions expressing how terms or concepts otherwise absent from the reducing theory T' are connected with those unique to the reduced theory T .

The requirement of supplying assumptions to connect the terms in T and T' presumes that scientific theories and their fundamental laws possess truth values. One broad approach to the nature of reduction, discussed by Nagel, denies this presumption. According to the *instrumentalist* account, theories and their laws are themselves neither true nor false; rather, they serve as rules or instruments through which one can infer certain experimental or observational claims from other such claims. On an instrumentalist reading, for example, the kinetic theory of gases is not a genuine description (of gases as systems of rapidly moving molecules), nor is it a true generalization of relations that hold among such objects; instead, it is a rule prescribing how one is to represent various observable states of affairs and how, given certain empirical data, one can infer other observational claims. As such, theories and their laws are not truths from which conclusions can be validly derived but functional devices or “inference tickets” in accordance with which claims may be drawn from other claims. Thus, the relevant issue for inhomogeneous reduction, on this account, is neither the cognitive content of the premises in a deductive explanation nor the logical or semantic relation of those premises to the conclusion. Rather, it is the relation between the sets of observational inferences licensed by the reduced and reducing theories. On the instrumentalist analysis of inhomogeneous reduction, T reduces to T' when all the observational claims that T licenses one to infer from some set of data are also licensed by T' , but not conversely.

NAGEL'S CORRESPONDENCE ANALYSIS OF INHOMOGENEOUS REDUCTION

While Nagel acknowledges that the instrumentalist analysis captures an important relation between the observational inferences licensed by theories when one reduces another, he denies that it gives an acceptable account of inhomogeneous reduction. Quite aside from the fact that scientists do in fact speak as if theories and laws are true or approximately true, the instrumentalist account fails to provide any unified story of how the theoretical concepts of one theory are related to those of another in a way that explains why various observational statements do (or do not) fall within the scope of reduced and reducing theories. Instead, Nagel proposes what he calls the *correspondence analysis of inhomogeneous reduction*. Like the instrumentalist proposal, it rests on certain crucial assumptions.

First, the correspondence analysis presumes that expressions belonging to some theory or domain of science have determinate meanings fixed by its own methods of explication. Of course, the meaning of certain terms in a theory or science may sometimes be specified with the aid of terms in some other theory or science, but it does not follow from this that the meaning of every term in a theory can be given by terms in some distinct theory. In particular, then, if the terms that are unique to the fundamental laws of a reduced theory T appear in the conclusion of a reductive explanation but not in the reducing theory T' itself, then those terms from T must have determinate meanings that are independent of T' . For example, *heat*, *temperature*, and *entropy* must have determinate senses specifiable by their role in thermodynamics, and these meanings must be explicable independently of statistical mechanics.

Second, the correspondence analysis presumes that observational terms shared by the fundamental laws of a reduced theory T and reducing theory T' possess meanings capable of being specified apart from T and T' —that is, they have meanings that are neutral with respect to whatever it is that distinguishes T and T' .⁴ Recall, for example, that thermodynamics and mechanics alike use the terms *volume* and *pressure*. The second assumption is that these terms possess a univocal sense that can be explicated independently of either T or T' .

Third, and finally, Nagel's correspondence analysis of inhomogeneous reduction presumes that reductions are a kind of deductive explanation.

BRIDGE LAWS

The logical-derivability condition of the correspondence analysis implies that, if a reduced theory T contains terms not appearing in the reducing theory T' , then T (or one or another fundamental law of T) can be the

conclusion of an explanatory deduction only if the reducing premises of T' are supplemented by some statement or statements expressing a connection between the vocabulary of T' and the terms unique to T . These statements, called *bridge laws*, are crucial to the success of inhomogeneous reduction in Nagel's explanatory-derivation model.⁵ If the ideal gas law (relating the temperature of a gas to its pressure and volume) is to be derived from the kinetic theory of gases (in which the concept of temperature makes no explicit appearance), then clearly some bridge law connecting the temperature of a system to an appropriate property of its constituent molecules must be added to the premises of the derivation.

But while the need for bridge laws in inhomogeneous reductions seems relatively clear, their status remains obscure. What kind of statements are bridge laws, and how should they be understood? One proposal is that they express logical or semantic connections between the meanings of the relevant reduced and reducing taxonomies. On this account, the bridge laws of inhomogeneous reductions are explicit definitions of the distinctive terms of a reduced theory T . But Nagel's first assumption—that terms unique to a reduced theory T have determinate meanings independent of the reducing theory T' —rules against this construal. Presumably, terms such as *heat* and *temperature* were meaningful before the advent of the kinetic theory, being fixed by the rules and habits of usage in thermodynamics; and, presumably, such terms can remain explicable by these procedures and retain their meanings, even if the laws in which they figure come to be reductively explained by molecular mechanics. These plausible assumptions would seem to be jeopardized if bridge laws were understood to express analytic connections, whereby the meanings of terms in the reduced theory T are logically tied to the meanings of terms in the reducing theory T' .

Nagel's own view about the status of bridge laws is that they express not logical but material (i.e., factual) connections between the states of affairs signified by the relevant terms in question. On this account, bridge laws are empirical hypotheses about the world. Nagel describes two forms that such connecting hypotheses may take: either they express relations between predicates and their extensions, or they express identities between objects or processes.

To understand Nagel's first alternative, it is helpful to employ the distinction between the extension of a predicate—the set of objects or processes to which the term applies—and the intension of a predicate—the universal property that the term connotes. (See "Extensions and Intensions" in the commentary on chapter 7.) Suppose that the reduced theory T contains a predicate, the term *blue*, connoting a property (blueness) that is not connoted by any of the predicates of the reducing theory T' . Nonetheless, T' might give conditions under which instances of blueness will arise. For example, T' might predict that light of a certain range of wavelengths is selectively transmitted by fluctuations in the density of

air in the earth's atmosphere, thus explaining why the sky is blue. In this example, the reducing theory *T'* does not contain the word *blue* but it provides a condition that is sufficient (though, in this case, not necessary) for the property of blueness to be manifested. The bridge law explicitly connects the predicate from *T'* (*selective transmission of light of a certain range of wavelengths*) with the predicate from *T* (*blue*) by saying that the former is sufficient for the latter. Nagel claims that bridge laws of this kind (involving either sufficient conditions, as in the present example, or necessary and sufficient conditions, as in other examples) are empirical hypotheses.⁶ He interprets these bridge laws as asserting that the extension of the reduced predicate is included in (or, in the case of necessary and sufficient conditions, coincides with) the extension of the reducing predicate. Thus, in our simple example, Nagel regards the bridge law as making the empirical claim that things that transmit light of a certain range of wavelengths are a subset of blue things.

Nagel's second alternative is that some bridge laws show that apparently distinct objects or processes are in fact identical. This type of bridge law would also be empirical. Just as it is an empirical discovery that the morning star is (i.e., is identical with) the evening star, so it is an empirical hypothesis that, say, light waves are (i.e., are identical with) electromagnetic waves of a certain range of frequencies, thus enabling the reduction of optics to electrodynamics. Bridge laws of this second kind do not simply specify conditions for the manifestation of properties connoted by predicates from the reduced theory; rather, they express an identity between entities initially reckoned as distinct.

8.2 | Feyerabend's Criticisms of Nagel's Model of Reduction

In "How to Be a Good Empiricist—A Plea for Tolerance in Matters Epistemological," Paul Feyerabend criticizes Nagel's model of intertheoretic reduction by arguing that it is inconsistent with an enlightened empiricist view of scientific change.⁷ Feyerabend's objections are part of a larger attack on the logical empiricist movement in twentieth-century philosophy of science.⁸ Although he is sympathetic with much of the empiricist tradition in philosophy, especially the empiricists' use of observation and experience to combat dogmatic metaphysics, Feyerabend thinks that modern philosophy of science has taken a wrong turn and has itself become dogmatic and a barrier to progress. He advocates a different brand of empiricism, a "good" empiricism that, unlike the "bad" empiricism of the logical empiricists, will discourage dogmatism and promote scientific progress. According to Feyerabend, a nondogmatic empiricism must welcome a plurality of theories, not for the purpose of advancing toward a final,

true theory that corresponds with “the facts,” but because a tolerance of alternative theories offers the best means of criticizing and evaluating the theories we currently accept. Thus, Feyerabend has two goals: to undermine Nagel’s reduction-based approach to scientific change and to defend a more liberal, theoretically pluralistic empiricist methodology.

We are already familiar with what Feyerabend calls a “cornerstone” of contemporary empiricism: a reductive explanation of T consists in the derivation of T from T' . We have also encountered two implications of this general account: first, that T' and any of its consequences within the domain of the reduced theory T must be consistent—i.e., logically compatible—with T , and, second, that either the descriptive terms of the reducing theory T' must be shared with those of the reduced theory T , or they must be explicitly connected to them via empirical bridge laws. These two implications may be stated as conditions that a new, more general theory must satisfy if it is to be accepted as explaining an existing theory in a given domain. According to the consistency condition, only those theories are admissible in a given domain that either contain the theories already used in that domain or are consistent with the theories in that domain. According to the condition of meaning invariance, meanings must not change with scientific progress; future theories must be expressed in such a way that they leave unaltered the claims made by the theories they explain.

In connection with his second goal—defending a more liberal methodology—Feyerabend is prepared to grant that many scientists and philosophers of science accept the consistency and meaning-invariance conditions as constraints on scientific progress. Postponing the question of their truth or falsity as descriptions of actual scientific change, Feyerabend wants to ask, Should they be accepted? Do they represent a reasonable empiricist methodology of science? Are they necessary conditions for scientific progress?

THE APPARENT UNREASONABLENESS OF CONSISTENCY AND MEANING INVARIANCE

According to Feyerabend, the consistency condition is unreasonably strong because it restricts the acceptance of new, alternative theories quite arbitrarily—on the basis of their relation to whatever theory happens to be accepted at the time. Suppose that a theory T is accepted as correctly accounting for the phenomena of some domain D , on the basis of a set F of confirming instances; and suppose a new theory T^* shares those confirming instances but contradicts T outside of F . On Feyerabend’s reading, the consistency condition recommends against accepting T^* simply on the basis of age and familiarity, not on the basis of any disagreement with the facts: had T^* rather than T been there first, the consistency condition would have worked against T rather than in its favor.

Feyerabend believes that considerations of this sort show that the consistency condition is not only unreasonable but also counterproductive. For suppose that an alternative (T^*) that shares the confirming instances of an accepted theory (T) cannot be eliminated by appeal to those factual instances. Defenders of the consistency condition insist that neither do those considerations recommend accepting T^* —in light of the empirical evidence, the alternative T^* is no better or worse than T . A genuine advance beyond T will arise, not from the invention of alternatives to T , but only from the emergence of fresh empirical facts, which by either supporting or requiring a modification of T , will bring about scientific progress. Feyerabend concedes that theories should not be changed unless some disagreement with the facts demands it. But why, asks Feyerabend, should we think that such facts exist and are available in the absence of alternatives to T ? According to Feyerabend, not only is the description of every fact dependent upon some theory or other, but many facts relevant to testing an existing theory are simply unavailable for discovery without considering alternatives to the existing theory.

To illustrate his claim that the discovery of new facts often depends on taking alternative theories seriously, Feyerabend gives the example of Brownian motion. Brownian motion is named after the English botanist Robert Brown (1773–1858) who first observed it in 1827. Using a microscope, Brown noticed that pollen grains suspended in water move about chaotically, rotating and darting from place to place in a random manner. Some scientists attributed Brownian motion to external vibrations disturbing the fluid, while others thought that it was caused by the action of light. This phenomenon is now taken to be a convincing proof of the kinetic theory of heat and a demonstration that the second law of thermodynamics is false. But for seventy years or more after Brown's discovery, its significance as a refutation of the second law was not appreciated because no one was able to give a convincing explanation of the phenomenon using the kinetic theory.⁹ When predictions based on the kinetic theory were tested, they failed to confirm the theory.¹⁰ Finally, Einstein in 1905, and the Polish physicist, Marian Smoluchowski (1872–1917) in 1906, were able to give a mathematical treatment of the phenomenon that made a number of testable, quantitative predictions. These predictions were confirmed by the French physical chemist Jean Perrin (1870–1942) and others.

Feyerabend interprets the history of Brownian motion in the following way:¹¹ Taken simply as an observed phenomenon, Brownian motion could not undermine the universal truth of the second law of thermodynamics.¹² Only when viewed from the perspective of the kinetic theory and statistical mechanics, and given an adequate mathematical analysis by Einstein and Smoluchowski, could the relevant facts of Brownian motion show that the second law is not universal but statistical.¹³ Since unearthing facts and recognizing their relevance to an accepted theory T is dependent upon

the existence of alternative theories as well confirmed as *T* but in disagreement with it, Feyerabend believes that the invention of alternatives is crucial to a properly empiricist methodology of science. Because the consistency condition proscribes theories in disagreement with *T*, it can only contribute to intolerance and dogma.

Feyerabend argues that the consistency condition also gives rise to a kind of self-deception about the success of current theories. By rejecting competitors to *T*, potentially refuting facts are suppressed. Consequently, judgments about the success of *T* are heightened, reinforcing even further the belief that investigating alternatives to *T* would be time ill spent. The alleged "success" of *T* is, thus, altogether fabricated, not a sign of its correspondence with nature, but simply of its staying power in the absence of empirical counterevidence. The accepted theory comes thereby to take on the character of myth, which by its very nature serves to perpetuate itself and infect how we think, talk, and explain. The result is conformism—a uniformity and unanimity of scientific belief born of self-deceit. Only by tolerating and encouraging variety of opinion can science avoid both the dogma and the self-deception common in many religious and political institutions.

Feyerabend extends his criticism of the consistency condition to an attack on meaning invariance. Tolerance of alternative theories brings with it a tolerance of alternative meanings, for changes in conceptual and theoretical systems produce changes in the interpretation of scientific terms. If our methodology permits mutually inconsistent, empirically adequate theories, then it must also permit mutually irreducible theories whose terms are unconnected by anything like factually correct bridge laws. This elasticity of meaning need not be occasioned by observational difficulties with an existing theory. For example, Feyerabend points out that the impetus theory of late scholastic and early modern science did not make any false predictions. But adopting Newton's theory of inertia produced better explanations of a wider range of phenomena. Such progress does not require that theoretical terms (such as *force* in the impetus and inertial theories) remain invariant in meaning when theories change. Since the reinterpretation of terms often accompanies theoretical change, the condition of meaning invariance is an unreasonable constraint on scientific methodology that would stifle progress.

In light of his rejection of the consistency and meaning-invariance conditions, Feyerabend notes three consequences of the methodology he defends. (See section 8 of his paper.) The first consequence is that we may relax our judgments about the propriety of metaphysical claims in formulating scientific theories. In the past, many empiricists have supposed that science must begin only with observable facts and proceed by generalizing from them, refusing to admit any intrusion of metaphysics. Contemporary empiricism allows metaphysical claims a role in new theories but with the proviso that confirmation of the theory will eventually enable

those claims to be purged. Feyerabend recommends an even more permissive attitude. On his view, metaphysical claims make up very primitive scientific theories. To this extent, even if they are inconsistent with accepted, well-confirmed, full-blown theories of nature, they serve precisely the critical and evaluative role that alternative theories should be encouraged to perform.

The second consequence of Feyerabend's methodology concerns the long-standing problem of how to justify inductive reasoning. We reason inductively when we make inferences from a sample to an entire population or from our past observations of some objects to a generalization about all objects of that type. Assume that a belief constitutes knowledge when it is true and justified. Since we must not change or give up what we know to be true, knowledge implies stability. Feyerabend claims that, since the methodology he endorses cannot allow such stability, the problem of justifying inductively formed beliefs is in fact a pseudoproblem: there is no such stability, hence no such knowledge, hence no such justification to be had.

Finally, the third consequence of Feyerabend's position is that so-called arguments from synonymy are likely to impede theoretical progress. Arguments from synonymy would evaluate a theory, not on the extent of its correspondence with the world, but on the extent to which its terms have the same meaning as (are synonymous with) the terms in some other theory that we already accept. Feyerabend gives the example of a neurophysiological theory that attempts to explain the mental life of persons solely in material terms. Proponents of synonymy will say that such a theory must fail, because it cannot provide synonyms for "pain," "thinking of Syracuse" or "feeling sad" in purely physical, neurochemical terms. Being in a certain kind of brain state, for example, does not mean the same thing as "thinking of Syracuse;" it is not what I mean when I say that I am thinking of Syracuse. Feyerabend asserts that "clearly, such criticism silently assumes the principle of meaning invariance" (943). Since Feyerabend rejects meaning invariance, he also rejects this kind of criticism as an unreasonable constraint on new theories: the terms they use should not have to mean exactly the same thing as terms already in use.¹⁴

In discussing this last consequence, Feyerabend endorses an even stronger result. He not only rejects synonymy; he also advocates irreducibility. By *irreducibility* Feyerabend means the absence of anything like Nagel's correspondence relation among theories. Now, it is one thing to claim that the conditions of consistency and meaning invariance are too restrictive; it is quite another to argue that in fact scientific change fails to meet the conditions of consistency and meaning invariance, that, in other words, Nagel's correspondence account of reduction is simply mistaken, and no theory is reducible to any other. We thus return to the first part of Feyerabend's paper, where he defends this latter, stronger claim of irreducibility.

THE APPARENT VIOLATION OF THE CONSISTENCY CONDITION

In our discussion of homogeneous reduction, we encountered the example of Galileo's law of falling bodies and Kepler's laws of planetary motion being said to reduce to Newtonian mechanics and gravitational theory. The reduction is homogeneous, since the fundamental concepts of those earlier-formulated laws (distance, velocity, acceleration, etc.) also appear in Newton's theory; moreover, the distinct and restricted domains to which Galileo's and Kepler's laws apply (terrestrial, celestial) are given a unified treatment by Newton's theory, from which Galileo's and Kepler's laws are said to be derivable. Feyerabend draws our attention to a point first made by Pierre Duhem (and repeated often by Karl Popper), namely, that the consistency condition required for such a derivation is apparently not met. (See section 4 of Pierre Duhem, "Physical Theory and Experiment" in chapter 3.) More precisely, the relevant consequences of Newton's theory are logically inconsistent with Galileo's law. Galileo's law asserts that the acceleration of a body in free fall near the earth's surface is a constant, yet an application of Newtonian theory to terrestrial free fall entails that the acceleration must vary with the distance of the body to the earth's center of mass. A similar result holds for Kepler's third law.¹⁵ What these and other historical cases show is that actual examples of scientific progress frequently violate the consistency condition. Furthermore, since the consistency condition is necessary for the success of a derivation, reduction is not a matter of deductive explanation at all. If anything is to be learned from such cases, it is not that a new, more general theory T' is one from which the fundamental laws of the earlier theory T can be derived as a special instance, but rather that T' simply differs from T and replaces it.

Nagel's response to this charge (in part I of the reading "Issues in the Logic of Reductive Explanations") is to acknowledge that the consequences derivable from a reducing theory T' may not always exactly coincide with the reduced laws of T but that this does not refute his model of reduction. It is a commonplace that newer, more comprehensive theories sometimes indicate the special conditions under which older, more restricted laws hold, in light of which some adjustment to the older laws may be necessary. Still, the earlier laws of T are generally close approximations to the consequences of T' . This being so, we can retain the deductive model of explanatory reduction at its core, taking homogeneous reduction to consist in the derivation of either the reduced laws or their close approximations from the reducing theory.

This result should not surprise us, for scientists often use simplifications and approximations when they derive laws from theories. For example, Nagel points out that in deriving the law for the period of a pendulum, we assume that the weight of the pendulum is located at a single point in the suspended bob, that the gravitational force on the bob is constant, and so on. Although such approximative assumptions are com-

mon, Nagel thinks that it would be absurd to say that scientists are mistaken in claiming to have derived laws from their theories. Given such assumptions, we count the laws as consequences of the theories; likewise, even if a reduced law of T is only a close approximation of what we actually derive from the reducing theory T' , it would be misleading to say that the law had not been reductively explained in the way Nagel's model describes.

THE APPARENT VIOLATION OF THE CONDITION OF MEANING INVARIANCE

Feyerabend argues that, like the consistency condition, the condition of meaning invariance is violated by actual scientific practice. His primary example is the relation of classical to relativistic mechanics and the concept of mass that appears in each theory. If we try to use the special theory of relativity to explain why classical, Newtonian, inertial mass is conserved in a closed system, the consistency condition again seems to be violated, for even if our two theories give experimentally indistinguishable results about the mass of objects at low velocities, it remains the case that, according to special relativity, the mass of the parts of our closed system—and so, too, its total mass—will vary according to the velocities of the parts relative to whatever coordinate system we use to conduct our measurements.¹⁶

Feyerabend extends this result to undermine the condition of meaning invariance, arguing that *mass* undergoes a change of meaning in the transition from classical to relativistic mechanics. For while it is true of classical mass, m_c , that the mass of an aggregate of parts is equal to the sum of the masses of the parts, this is not so for relativistic mass, m_r , where the mass balance is a function of the relative velocities of the parts. Moreover, while classical mass m_c is a (one-place) *property* of an object, independent of coordinate system, relativistic mass m_r is a (two-place) *relation*, involving relative velocities, between an object and a coordinate system. In light of this, Feyerabend claims that we must conclude that *mass* as it appears in classical mechanics and *mass* as it figures in relativity theory do not have the same meaning. As before, he contends that the successor theory T' has not absorbed or incorporated T , as Nagel's account requires, but straightforwardly differs from T and replaces it. As Feyerabend says elsewhere:

What happens when transition is made from a restricted theory T' to a wider theory T (which is capable of covering all the phenomena which have been covered by T') is something much more radical than incorporation of the *unchanged* theory T' into the wider context of T . What happens is rather a *complete replacement* of the ontology of T' by the ontology of T , and a cor-

responding change in the meanings of all descriptive terms of T' (provided these terms are still employed).¹⁷

(Contrary to the notation used in this commentary and in Nagel, Feyerabend uses T' to stand for the reduced theory and T to stand for the reducing theory.)

It might be objected that Feyerabend's examples of meaning change fail to show anything so strong as this. At most they show that actual scientific practice sometimes violates the condition of meaning invariance for highly theoretical terms (such as *mass*), leaving the meaning of the observational terms shared by T and T' unchanged. But Feyerabend intends his arguments to apply to observational terms as well. The significance of denying meaning invariance for observational terms is best judged against the backdrop of a brief account of the distinction between observational and theoretical terms.

One way of construing the theoretical-observational distinction is to see it as marking off the descriptive predicates that appear in observational (or experimental) laws from the predicates that appear in theoretical laws. Observational laws are generalizations about phenomena accessible to the senses; accordingly, they are expressed in the language of observation and refer to perceived or perceivable objects, processes, properties, and so on. Theoretical laws, on the other hand, are not generalizations about observable phenomena but rather help to give unified explanations of observable phenomena and their laws. Since theoretical laws typically make reference to unobservable things and properties, they are expressed in a more abstract language. Statements of the freezing and boiling points of water as a function of pressure are observational laws, and *freezing point*, *boiling point*, *temperature*, and *pressure* are observational terms. A law expressing the degrees of rotational freedom for helium molecules is a theoretical law, and *degree of rotational freedom* and *helium molecule* are theoretical terms. (For an extended account of the theoretical-observational distinction, see chapter 9, below.)

To appreciate the force of Feyerabend's denial of meaning invariance for observational as well as theoretical terms, consider the empiricist view (held by Nagel and others) of how observational and theoretical terms get their meaning. The meaning of both sorts of terms is supposed to come from one source: experience. Observational terms get their meaning directly from observation and from test procedures that can be applied independently of any theoretical context. So even though theories change, the meaning of observational terms remains constant. Theoretical terms also get their meaning from experience, but indirectly: the meaning of a theoretical term is supposed to "seep up" from the experimental laws derivable from the theory in which the term occurs. So, in general, the meaning of theoretical terms changes as theories change, but if rival theories predict many of the same experimental laws, there will be a consid-

erable overlap in the empirical meanings of the theoretical terms involved. Thus, observation terms can be used to state experimental laws that competing theories then seek to predict and explain. Moreover, because observation terms are independent of any theory, those experimental laws can serve as the independent basis for evaluating the explanatory and predictive success of those theories. By denying meaning invariance both to theoretical and observational terms, Feyerabend effectively cuts off *any* shared meanings between earlier and later theories and denies to experimental laws their role as neutral, theory-independent arbiters of theoretical success.

What leads Feyerabend to deny meaning invariance to observational terms? On his account, the content of such terms is not exhausted by observational procedures but also depends on our theories about the nature of the things we are observing. Thus, even if our observational procedures for ascertaining some perceivable aspect of an object do not change from the use of T to the adoption of T' , nevertheless T and T' may diverge in how they represent the nature of that aspect of the object—as a property, as a relation, as a process, and so on. And this divergence will affect the meaning of our observational terms. If Feyerabend is correct about this, then the traditional distinction between the meaning of observational and theoretical terms is jeopardized.

Feyerabend's primary example of meaning variance is Einstein's special theory of relativity. (Kuhn also uses the same example in his "The Nature and Necessity of Scientific Revolutions," reprinted in chapter 2.) Feyerabend claims that the meaning of the theoretical term *length* is different in classical mechanics than it is in special relativity. If this is true, then classical mechanics is not reductively absorbed by Einstein's theory but is completely replaced by a wholly different theory. Specifically, Feyerabend denies that we can assign any common empirical meaning to the term *length* in the two theories by connecting it to a neutral observation language. For, he argues, any such observation language will, at best, yield a representation of length contraction that corresponds to Lorentz's electromagnetic theory, not to Einstein's relativity theory.¹⁸

Understanding Feyerabend's claim requires that we say something about the difference between Einstein's theory and Lorentz's. Both theories yield the same set of equations (the Lorentz transformations) and both predict the length contraction of moving objects. But the interpretation of length contraction in the two theories is radically different. Lorentz's theory is a classical theory in which any contraction of a physical object must be due to a physical force compressing the object and making it shorter. Lorentz (following Fitzgerald) hypothesized that all objects are composed of electrons and molecules and that all the forces that hold matter together are electrical. When an object moves through the electromagnetic aether (or, in other words, when an object moves with respect to absolute space), these forces become stronger, causing the object to shrink. Thus, on Lor-

entz's theory, the real length of an object is the length it has when it is at rest with respect to absolute space, and length contraction is a physical effect caused by the motion of electrons in absolute space.

Einstein's theory is totally different from Lorentz's. Einstein bases his theory on the empirical fact that the speed of light is the same in all frames of reference, together with an operational analysis of the concepts of length, time, and simultaneity. His derivation of the Lorentz transformations involves no hypotheses about the structure of matter and it dispenses with the notion of absolute space. On Einstein's theory, all objects moving with constant velocity with respect to an inertial frame of reference must appear shorter when measured in that frame of reference, regardless of their composition. Thus, in Einstein's theory there is no absolute space and no physical forces causing length contraction.

Feyerabend's point is that our ordinary, commonsense views about objects are embedded in our observational language. For example, Feyerabend presumes that it is part of what we ordinarily mean by observational terms such as *physical object* and *length* that physical objects will shrink only when a physical force compresses them. If this is true, then there is no way that we can interpret the theoretical terms in Einstein's special theory of relativity using such an observational language. We may assign empirical meanings to the terms in the Lorentz-Fitzgerald transformation equations, but the interpretation we give will be Lorentz's interpretation, not Einstein's. In brief, Einstein's theory involves a radical reinterpretation of our ordinary concepts of length and time. Thus, we should not expect that an observational language in which the classical, Newtonian concepts of length and time are embedded will be able to assign empirical meaning to the theoretical terms in Einstein's theory.

According to Feyerabend, even the most innocent-seeming observational language is theory dependent, and each theory determines its own distinctive set of observational statements. Thus, he concludes that the traditional empiricist distinction between observational and theoretical terms fails to give a satisfactory account of meaning in scientific theories and that the standard reductive picture of the relation between theories fails along with it.

Nagel's reply to the "replacement" account of intertheoretic reduction (part 3 of his "Issues in the Logic of Reductive Explanations") begins with a summary of Feyerabend's position. Since the meaning of every term—observational and theoretical alike—depends upon the particular theory in which it occurs, there is no theory-independent observational core to ground the interpretation of any sentences making up a theory; rather, the theory itself is meaningful independent of observation. Accordingly, a change from one theory to another brings with it a change in the meanings of all terms, and hence no theory *T* can be reductively derived as an explanatory consequence of *T'*.

The standard empiricist view is that scientists should formulate the-

ories adequate to the facts of natural phenomena as given by theory-neutral observation. Suppose Feyerabend's view to the contrary is correct. That is, suppose a transition is made from some existing theory *T* about a domain *D* of phenomena to a distinct theory *T'*, and suppose that the vocabularies of *T* and *T'* share no meanings in common. Nagel's first objection has two parts. First, if Feyerabend's account is correct, then no sense can be made of the claim that *T* and *T'* are either consistent or inconsistent (compatible or incompatible), since consistency (and compatibility) are logical relations obtaining between statements whose terms have shared meanings. (If you assert "*x* is *F*" and I assert "*x* is not *F*," we have not made inconsistent claims unless you mean by "*F*" what I mean by "*F*".) Second, if the meaning of observation terms in *T* is a function of *T* itself, then any laws in which those terms figure cannot be at odds with the observational claims of any alternative theory relative to which we may wish to assess the empirical adequacy of *T*. If there are no objective, theory-independent grounds for fixing the meaning of any terms in a theory, every theory is equally immune from criticism by adherents to every other theory.

Nagel's second objection is a further attack on the intelligibility of Feyerabend's position. Feyerabend tells us that when a transition is made from an existing theory *T* to a distinct theory *T'* "capable of covering all the phenomena that have been covered by *T*" (916) *T* and *T'* share no meanings in common. But how, Nagel asks, in the absence of common meanings, can we compare the relative scopes of *T* and *T'* and their degree of overlap? If the descriptive means by which one is able to circumscribe the relevant phenomena are theory laden, and thus undergo meaning changes in the transition from *T* to *T'*, it is unclear what neutral account is available for expressing the notion that *T'* is capable of covering all the phenomena that have been covered by *T*. (Expressing this notion is no difficulty for Nagel's correspondence account, since bridge laws are understood as empirical hypotheses about how the extensions of the terms of *T* are related to those of the terms of *T'*.)

There is a further issue concerning Feyerabend's thesis of meaning variance. In chapter 7, we encountered the distinction between a term's intension and its extension—or, as it is nowadays more commonly expressed, between the sense and the reference of a term. To which of these traditional aspects of meaning is Feyerabend's denial meant to apply? Reasons (however good or bad they may be) for claiming that, say, *mass* as it figures in classical mechanics and *mass* as it appears in relativity theory have different meanings (i.e., have different senses) are not yet reasons for believing that each of these homophones diverge in reference: classical mechanics and relativity theory may, for all that, refer to a single physical magnitude, diverging on the properties their hypotheses ascribe to it. Thus, even if one endorses a position according to which senses of terms are closely tied to the theories in which they figure, we may yet have grounds

for believing, as Nagel's remarks imply, that there is an objective and common domain of natural phenomena to which both T and T' may be said to apply.

We can summarize the contrast between Nagel's and Feyerabend's views on meaning by returning to Feyerabend's criticism of synonymy as a constraint on new theories. (See pages 939–41) Feyerabend protests that synonymy—the requirement that the meaning of terms in established theories not change when new theories are introduced—would stifle innovation. He reaches this conclusion by adopting a holist theory of meaning according to which any new claim about, say, temperature or mass, automatically changes what the terms *temperature* and *mass* mean. As we have seen, Nagel rejects the holist theory of meaning because it would render impossible any comparison between theories (including the ones that Feyerabend makes). In its place, Nagel relies on the traditional distinction between a term's intension (sense) and its extension (reference). Even if the intension of a term varies with theoretical context, its extension—the things or processes that the term refers to—can remain the same. In this way, contra Feyerabend, rival theories can make competing claims about the same objects.

Nagel's account of reduction leaves open the possibility that two theories might use the same term not only with different intensions but also with different extensions. In such a case, any reductive relation between the theories must be inhomogeneous, secured by bridge laws. To illustrate this possibility, let us return to the disputed example of classical mechanics and relativity theory and allow, for the sake of argument, that the extension of the term *mass* is not the same in both theories. Nagel would insist that we introduce two new terms that are typographically distinct, m_c and m_r , say, and then try to express the relation between the extensions of those terms in a bridge law. Remember that the bridge law is not a definition or a statement of synonymy. Rather, it asserts an empirical hypothesis about how the extension of m_c is related to the extension of m_r . Whether or not such a bridge law can be found in this case is a matter of dispute. (See the discussion of Nickles, which follows in this commentary.) Nonetheless, the point remains that, contrary to Feyerabend, Nagel's condition of meaning invariance is not a barrier to theoretical innovation once it is understood as Nagel intended, namely, as a prohibition of referential ambiguity when theories are compared.

8.3 | Nickles on Two Concepts of Intertheoretic Reduction

What assessment should we make of our discussion to this point? On the one hand, the attractiveness of Nagel's account is hard to deny: surely

there is *something* right in the view that, say, claims about temperature can be reductively explained in terms of the mean kinetic energy of molecules, that the laws of some theory *T* about a certain domain can be consolidated within a distinct theory *T'* about some wider domain, and that such laws will often be deductive consequences of the more general theory. On the other hand, it is scarcely deniable that succeeding theories sometimes contradict their predecessors and that scientific change can bring with it an evolution in our concepts.

Until now our discussion has proceeded under the implicit assumption that reduction is a single phenomenon and that the central task is one of determining which model—instrumental, correspondence, replacement, or something else—best captures this relation between theories. In other words, we have been supposing that the relations between thermodynamics and statistical mechanics, between Galilean and Newtonian laws of motion, between classical mechanics and relativity theory, and so on each alike instantiate some one concept of intertheoretic reduction. But perhaps the possible relations among theories are so diverse that no single model can do justice to all these examples of reduction. This is the view taken by Thomas Nickles in “Two Concepts of Intertheoretic Reduction.”

From the many examples of reduction offered by Nagel and others, Nickles believes that we can form a relatively clear concept of a reduction relation connecting narrower, less fundamental theories with broader, more fundamental ones. In such cases as the reduction of thermodynamics to statistical mechanics, or physical optics to electromagnetic theory, the basic claims of a putatively autonomous theory are consolidated into or absorbed by the theoretical scheme of an overlapping but richer theory. This achievement of conceptual and ontological economy is obtained chiefly by the derivational explanation Nagel has proposed.

But Nickles argues that not all reductions are of this sort. In the case of classical mechanics and relativity theory, for example, the broader, more general theory of relativity reduces to the narrower, more restricted classical mechanics in the limit of low velocities (as velocities approach zero). We have encountered this sort of reduction in the connection between classical and relativistic mass: the relativistic measure of mass reduces to the classical measure in the limit. Nickles makes the same point about the relativistic and classical definitions of linear momentum. What we have in these and other similar cases, in which successor theories reduce to their predecessors under limiting transformations, is an inversion of the standard concept of reduction sketched above. But these examples, no less than those above, are among our paradigm examples of reduction. Since we cannot claim both that less general theories reduce to more general ones and that more general theories reduce to less general ones in the same sense of *reduce*, we must grant that *reduce to* expresses different relations in different contexts. In short, there must be more than one kind of reduction.

Nickles labels the first sort of reduction, whereby T is consolidated by T' to produce a conceptually and ontologically more efficient system, *reduction*₁. Since reduction₁ is a relation between consistent (and typically concurrent) theories in different domains, it is said to be a domain-combining reduction. The second kind of reduction, called *reduction*₂, is primarily a relation between T and its historical successor T' and is what Nickles describes as a domain-preserving reduction. The distinction between these two intertheoretic relations can be clarified by noting several further important differences between them.

First, in exhibiting how the fundamental claims of T are superfluous or derivative vis-à-vis the more general theory T' , reduction₁ is a derivational reduction of T from T' . It is in this way that thermodynamics and statistical mechanics are united by a domain-combining reduction₁: claims about temperature and entropy being deductively explained in molecular-kinematic terms. But reduction₂ is neither an ontological nor an explanatory reduction as normally construed. The relation of T' to T in such cases is primarily heuristic and justificatory. It is heuristic insofar as the formulation of new theories can plausibly be guided by the idea that, other things being equal, new theories will yield accepted results as a special case. The relation is justificatory in the sense that a successful reduction₂ shows that T' can give an adequate account of the phenomena treated by T .

Second, since a successor theory, T' , may approach its predecessor, T , in the limit in more than one way, T' may reduce₂ to T in several respects, but given the nature of explanatory and ontological reduction, T can reduce₁ to T' in only one way.

Finally, while reduction₁—being a relation of explanatory derivation—must hold between compatible theories, this is not the case for reduction₂. The special theory of relativity is a replacement of classical mechanics, with which it is, strictly speaking, inconsistent. And given that the function of reduction₂ is primarily heuristic and justificational, Nickles claims that it is less sensitive to meaning change than reduction₁ can be.

REEVALUATING NAGEL'S ACCOUNT

Nickles's proposal gives us a helpful perspective from which to reevaluate Nagel's account of reduction. Domain-combining reductions₁, between theories that were regarded historically as being in distinct domains, will generally be inhomogeneous—Nickles calls them *heterogeneous*—and will involve the use of bridge laws to connect distinctive vocabularies. Through the use of such bridge laws, the domain-combining absorption of T by T' is effected by the derivation of T from the compatible but more general theory T' . Domain-preserving reductions₂ of a successor T' to its predecessor T will typically be homogeneous but never derivational. Thus, according to Nickles's account, Nagel is best viewed as capturing only

domain-combining reduction,₁ and, in treating all reduction as derivational, casting such intertheoretic relations in too narrow a mold. By doing so, Nagel opens his account to Feyerabend's (and Kuhn's) objections against the meaning-invariance and consistency conditions. Let us return briefly to these.

While we have noted several misgivings about Feyerabend's rejection of meaning invariance, Nickles himself does not explicitly criticize this rejection. But he does criticize one attempt to deflect the force of this rejection in the case of homogeneous reductions by claiming that the attempt concedes too much to Nagel's critics. Nickles's point is that if we grant that the homophonic terms of T and T' differ in meaning, we might simply consider the reduction as heterogeneous (or inhomogeneous), connecting m_c with m , by appropriate bridge laws, allowing at once for the claim that T and T' diverge in their descriptive vocabularies but nevertheless providing for derivational reduction. But by thus treating all reduction as in the end heterogeneous, we undermine what seems correct about distinguishing heterogeneous from homogeneous cases. That is, we effectively undermine our ability to distinguish domain-preserving reductions, from domain-combining reductions;₁ we would have to take the classical-relativistic reduction to be one of conceptual and ontological consolidation by explanatory derivation, on a par with the reduction of thermodynamics to statistical mechanics or optics to electromagnetic theory. And that, Nickles thinks, is to give up too much. Although he does not say so explicitly, one is left with the impression that he believes this gives up too much in part because it amounts to granting the denial of meaning invariance across the board, to all theoretical change, even in cases of heterogeneous, domain-combining reductions.₁ But even if *mass* changes in meaning from classical to relativistic theories, surely *pressure* and *volume* are univocal from thermodynamics to statistical mechanics.

Nickles's appraisal of the consistency condition is similar. The objection that reduced and reducing theories are always or nearly always incompatible is too strong. For many logical incompatibilities are simply the consequence of historical contingencies. Nickles argues that it would be nothing shy of miraculous if a theory T , arising from one line of research (say, traditional thermodynamics or physical optics) were, without any adjustments of detail whatsoever, completely consistent and derivable from some distinct theory T' , arising from a distinct line of research (such as statistical mechanics or electromagnetic theory). Often the adjustments of detail in T can be accomplished within the context of its original historical framework, and in any case there is no conceptual or logical requirement that such divergence of detail must always arise. Moreover, Nagel's critics have not been properly mindful of the fact that philosophers (and sometimes scientists) operate with deliberately idealized models of theories and that such models often depart from fully accurate representations in ways relevant to supposed inconsistencies between T and T' . Nevertheless,

while Nickles rejects the claim that all reductions₁ must fail because of incompatibilities between T and T' , many reductions do fail to meet the consistency condition. Prominent among these are domain-preserving reductions₂ of successor theories to their predecessors. We should not pretend that relativity theory is consistent with classical mechanics. But neither should we conclude, with Feyerabend, that there is no reduction to be found there at all. Instead, we should grant that some kinds of intertheoretic reduction depart from Nagel's derivational model. Thus, reduction₂ is immune to Feyerabend's consistency objection.

The existence of reduction₂ shows that Nagel's model does not express *necessary* conditions for (all kinds of) intertheoretic reduction. Other examples seem to show that neither are Nagel's conditions *sufficient* for reduction. It is typically presumed that the molecular-kinetic theory of matter includes, in addition to a class of statistical assumptions, the fundamental laws of mechanics. If this is correct, then the presence of mechanical laws in kinetic theory entails, on Nagel's account, that mechanics trivially reduces, by derivation, to the kinetic theory. But surely this result is not to be explained as a realization that, contrary to our best intuitions, mechanics is less fundamental or less basic than the kinetic theory and reduces to it. Rather, this result shows that Nagel's account cannot distinguish the reduction of one theory to another from the application of one theory to special domains. Meeting the conditions of Nagel's derivational account is not sufficient for reduction. A similar result (first noted by Lawrence Sklar) emerges from the Wiedemann-Franz law, which one can use to deduce the thermal conductivity of a metal from its electrical conductivity.¹⁹ No one has ever suggested that the theory of heat conduction is reducible to the theory of electrical conduction. One way of expressing the lesson of this example is that the existence of laws that correlate properties in bridge hypotheses do not suffice for effecting a reduction.

Nickles considers the proposal, recommended by Sklar, that we might improve the standard correspondence account of reduction by replacing bridge laws that merely correlate entities or properties with bridge laws that identify them. (Recall that Nagel himself suggested that some bridge laws have the form of identity statements.) This proposal is not without its problems. As Nickles argues, such an account would still seem to classify as reductions many intertheoretic relations that clearly are not reductions at all—as when the identification of light beams with electromagnetic radiation of a certain frequency supports a partial “reduction” of electromagnetic theory to geometric optics. Moreover, it is often unclear exactly what identifications a reduction would support. This is especially so in the reduction of nonstatistical theories such as thermodynamics or classical mechanics to statistical theories like statistical mechanics or quantum theory, where there are alternative formulations of the reducing theory's concepts. Distinct formulations may offer different heuristic or metatheoretical advantages. Indeed, as Nickles goes on to suggest, some formulations may

weigh against identification altogether in favor of much weaker correspondence rules. For example, while the classical concept of particle position most closely resembles the quantum-theoretical concept of position, it proves far more illuminating to correlate classical position with a statistical average of the values obtained when position is measured in an ensemble of systems, all of which are initially prepared in the same quantum state. In this and many similar cases in quantum mechanics, the correspondence of classical quantities with expectation values does much to justify quantum theory as a deep account of the phenomena classical mechanics set out to explain. As before, Nickles's argument is that in many paradigmatic examples of reduction, the heuristic and justificational advantages of a theory T' relative to its predecessor T overshadow any efforts to see the relation between the theories as one of ontological identification or derivation via bridge laws. In such cases, reduction₂ is the most fitting model for intertheoretic reduction.

One problem with Nagel's view about the status of bridge laws is the problem of symmetry: when bridge laws are expressed as biconditionals (which is a symmetric relation), how best can we explain the intuition that T reduces to T' , but T' does not in that same sense reduce to T ? In the final section of his paper, Nickles considers this issue as it applies to reduction₂. Supposing that the mapping of one theory into or onto another via some limiting transformation effects a reduction₂, the threat of symmetry would seem to arise only if the inverse of the limiting transformation exists, which typically it does not.²⁰ Nonetheless, the symmetry problem can still arise if there is a limiting operation that maps T' into T and another operation (not the inverse of the first) that maps T into T' . Nickles illustrates this possibility with the following fictional example of two little theories (or parts of theories), T and T' ,

$$T: w = ax + 2y + g \quad T': z = bx + ey + d$$

where w , x , y , and z are variables and a , b , d , e , and g are numerical constants. Suppose first that $a = b$, and $e = 2$. In this case T and T' differ only with respect to g and d . Suppose now that T is an old, predecessor theory, with $g = 0$, and suppose that T' is a new theory, with $0 \neq d \ll 1$. If it is legitimate to say that T' reduces₂ to T in the limit as $d \rightarrow 0$, is it not equally legitimate to say that T reduces₂ to T' in the limit as $g \rightarrow d$? And yet, intuitively, it seems that if one theory reduces to another in the limit, then the former is in some respect a generalization of the latter, which is an asymmetric relation among theories.

Nickles's response to the symmetry problem is to prohibit unrestricted changes in the values of a theory's numerical constants. Indeed, if the values of variables and constants alike are allowed to approach different magnitudes, any equation could be reduced to any other. Changes in the values of constants are permitted in reductive transformations only when

weigh against identification altogether in favor of much weaker correspondence rules. For example, while the classical concept of particle position most closely resembles the quantum-theoretical concept of position, it proves far more illuminating to correlate classical position with a statistical average of the values obtained when position is measured in an ensemble of systems, all of which are initially prepared in the same quantum state. In this and many similar cases in quantum mechanics, the correspondence of classical quantities with expectation values does much to justify quantum theory as a deep account of the phenomena classical mechanics set out to explain. As before, Nickles's argument is that in many paradigmatic examples of reduction, the heuristic and justificational advantages of a theory T' relative to its predecessor T overshadow any efforts to see the relation between the theories as one of ontological identification or derivation via bridge laws. In such cases, reduction, is the most fitting model for intertheoretic reduction.

One problem with Nagel's view about the status of bridge laws is the problem of symmetry: when bridge laws are expressed as biconditionals (which is a symmetric relation), how best can we explain the intuition that T reduces to T' , but T' does not in that same sense reduce to T ? In the final section of his paper, Nickles considers this issue as it applies to reduction₂. Supposing that the mapping of one theory into or onto another via some limiting transformation effects a reduction₂, the threat of symmetry would seem to arise only if the inverse of the limiting transformation exists, which typically it does not.²⁰ Nonetheless, the symmetry problem can still arise if there is a limiting operation that maps T' into T and another operation (not the inverse of the first) that maps T into T' . Nickles illustrates this possibility with the following fictional example of two little theories (or parts of theories), T and T' ,

$$T: w = ax + 2y + g \quad T': z = bx + ey + d$$

where w , x , y , and z are variables and a , b , d , e , and g are numerical constants. Suppose first that $a = b$, and $e = 2$. In this case T and T' differ only with respect to g and d . Suppose now that T is an old, predecessor theory, with $g = 0$, and suppose that T' is a new theory, with $0 \neq d \ll 1$. If it is legitimate to say that T' reduces₂ to T in the limit as $d \rightarrow 0$, is it not equally legitimate to say that T reduces₁ to T' in the limit as $g \rightarrow d$? And yet, intuitively, it seems that if one theory reduces to another in the limit, then the former is in some respect a generalization of the latter, which is an asymmetric relation among theories.

Nickles's response to the symmetry problem is to prohibit unrestricted changes in the values of a theory's numerical constants. Indeed, if the values of variables and constants alike are allowed to approach different magnitudes, any equation could be reduced to any other. Changes in the values of constants are permitted in reductive transformations only when

they make physical sense given the theories involved. Although he is unprepared to offer any general formula specifying constraints that all legitimate reductions₂ must meet, Nickles concludes his paper by proposing several conditions to help mark out genuine reductions₂. First, reductively related theories must apply to the same or significantly overlapping phenomena. Second, entire sets of equations must be reduced₂ to others. Third, as we have discussed, transformations, especially those involving constants, must make physical sense. These requirements will severely restrict trivializations similar to the one envisioned above. Finally, reduction₂ will be taken as a relation of new to previously established theories, with a reducing₂ theory having to account for the intelligibility of its predecessor. In this way the empirical content of established predecessors will impose constraints on the variables and their relations in the successor theory.

8.4 | Kitcher on Reduction, Classical Genetics, and Molecular Biology

Nagel and others have proposed models of the reduction relation that is supposed to hold among theories when scientific progress occurs. As we have seen, some philosophers such as Feyerabend are skeptical about the existence (and methodological propriety) of any such reduction relation among scientific theories. Others, such as Nickles, are prepared to recognize reduction but argue that it is considerably more complicated and less unified than Nagel envisioned. Thus, the philosopher of science who remains sympathetic to the idea of reduction is left to decide in what historical cases, and to what extent, some form of reduction has taken place. Typically, the examples used by philosophers of science are drawn from the physical sciences, usually physics and chemistry. But the life sciences should not be ignored. They, too, provide important cases of scientific progress in which the philosopher of science has to judge whether, and to what extent, one theory has been reduced to another.

In his paper “1953 and All That: A Tale of Two Sciences,” Philip Kitcher presents some new ideas on the nature of theories and the relations among them, by asking whether or not two central theories in biology—classical genetics and molecular biology—stand in a reductive relation to one another. Classical genetics is the familiar textbook account of genetics developed in the first decades of the twentieth century by T. H. Morgan and his associates through their research on the fruit fly *Drosophila*. Morgan’s work extended and refined Mendel’s pioneering factor theory of heredity by identifying chromosomes as the carriers of Mendel’s factors. These factors were soon renamed *genes*. Molecular genetics began with Watson and Crick’s discovery of the structure of DNA in 1953.

In addition to expanding our discussion of intertheoretic reduction to include the life sciences, Kitcher's paper illustrates how philosophers of science strive to improve our understanding of intertheoretic relations by looking in some detail at particular scientific theories. Among philosophers of biology, a relatively common position has been that classical genetics reduces to molecular genetics. Kitcher rejects this position. But before taking up his arguments against reductionism in genetics, let us briefly review some of the central notions at work in classical and molecular biology.

Classical genetics regards genes as segments of chromosomes, each having its relative position or locus on the string of genes making up the chromosome. The chromosomes of diploid organisms are paired, members of these matched pairs being called *homologous chromosomes*. The genes that can occur at some locus on a chromosome are called *alleles*. If different alleles occur at the same locus on a pair of homologous chromosomes, the organism is said to be *heterozygous at that locus*.

There are two different kinds of cell division: mitosis and meiosis. When a cell divides by mitosis, it produces two cells that are genetically identical to the original cell; each pair of chromosomes is duplicated exactly. Mitosis is the cloning process by which multicellular organisms grow and many single-celled organisms reproduce. Unlike mitosis, meiosis involves a reduction in the number of chromosomes and is the process by which sex cells (gametes) are formed in the ovaries and testes. At meiosis, the pairs of chromosomes line up, and each member separates from its partner to form two gametes, each containing half the required number of chromosomes. Fertilization occurs when the male and female gametes unite to yield a cell with a full chromosomal complement.

Although the new individual that results from fertilization receives half its chromosomes from each parent, the paternal and maternal genes in homologous pairs of chromosomes do not remain separate for very long. Through several mechanisms they soon become mixed up together and arranged in a different order. For example, homologous chromosomes often wrap around each other, break, and then recombine to change the arrangement of genes. By such crossing over and recombination, chromosomes come to contain both maternal and paternal genes. The likelihood and frequency with which some gene comes to have new neighbors on the chromosome through recombination is a function of that gene's proximity to others and to the site of crossover. The organism's genetic makeup or genotype together with environmental effects give rise to the organism's outward, or phenotypic, characteristics. Classical genetics is thus a study of hereditary phenomena, explaining phenotypic characteristics in terms of gene-chromosome relations, gene loci and the alleles at these loci, gene transmission in gamete and zygote production, the dominance and recessiveness of genes, and so on.

The scope of molecular biology is somewhat different. It is, in effect,

a study of biosynthesis at the molecular level, a theory about the molecular structure of fundamental genetic units, how they are formed and modified, and how, above all, these genetic units produce the proteins that ultimately determine phenotypic characteristics. The molecular analogue to the classical Mendelian gene is a segment of deoxyribonucleic acid (DNA)—a long spiral-shaped molecule composed of two strands of sugar and phosphate molecules connected by two pairs of complementary bases (adenine and thymine, guanine and cytosine).

The DNA in genes has two main functions: replication and transcription. In gene replication, the strands of DNA split apart, each strand serving as the template for the formation of two new DNA molecules. In transcription, the DNA serves in the production of a similar molecule called *ribonucleic acid* (RNA). RNA, whose single-strand sequence is uniquely determined by the DNA sequence, is responsible for the synthesis of proteins in ribosomes, through a process of translation. Proteins are synthesized by stringing together amino acids in long chains. By virtue of their particular sequence of amino acids, proteins take on complex three-dimensional secondary structures that enable them to perform their many functions: regulating the transcription process itself, producing other proteins via RNA, and serving as enzymatic substrates for biosynthetic pathways. Thus, the molecular structure of DNA—that is to say, the sequence of its base pairs—uniquely determines (in concert with the environment) all phenotypic characteristics of an organism. By deciphering the genetic code, molecular biologists discovered which triplets of DNA base pairs code for which amino acids and thus which sequences of base pairs (genes) produce which proteins.

THE APPARENT FAILURE OF REDUCTIONISM IN GENETICS

Kitcher approaches the issue of reduction in genetics²¹ by allowing that some core, received account of intertheoretic reduction is able to capture many of the standard examples of this phenomenon in the physical sciences. Reductionists claim that the relation of classical to molecular genetics is sufficiently similar to these examples to count as yet another instance of this reductive phenomenon. Given the standard model of reduction as a derivational explanation by T' of reduced laws of T with the help of bridge laws connecting the distinctive vocabularies of T and T' , the reductionist position requires us to accept three fundamental theses:

- R1 Classical genetics contains general laws about the transmission of genes which can serve as the conclusions of reductive derivations.
- R2 The distinctive vocabulary of classical genetics (predicates like “① is a gene,” and “① is dominant with respect to ②”) can be linked to the vocabulary of molecular genetics by bridge principles.

R3 A derivation of general principles about the transmission of genes from principles of molecular biology would explain why the laws of gene transmission hold (to the extent that they do.) (973)

In the first part of his paper, Kitcher argues against reductionism, claiming that (R1), (R2), and (R3) are false. Let us consider these in turn.

(R1) As we noted earlier in this commentary, philosophers of science commonly identify theories with some manageable set of fundamental laws. Kitcher's first observation is that classical genetics is not presented by philosophers of science in this way at all (even by reductionists). Rather, it is presented by reference to the particular subject matter and to particular methods of investigation. The explanation for this fact is that nothing approaching genuine laws can be found in the research papers and textbooks themselves. It is true that, before Morgan's research in the early decades of this century, geneticists did make reference to "Mendel's laws." But these "laws" have since been proved false.

Kitcher's second point is taken from a particular example of the failure of these so-called laws. Mendel's second law claims that genes at different loci in diploid organisms are transmitted independently when gametes are formed by meiotic division. But since genes are segments of chromosomes, alleles on the same chromosome will tend to be transmitted together. In response to the reductionist claim that such mistakes in the laws can be corrected, Kitcher points out that it is not so much that the law is false as that the law itself plays no role in classical genetic research. Classical genetic researchers use cytological techniques of a quite general sort to see what can be learned from them. So any change these researchers might make to the second law would be simply a special-case result of the general cytological methods, with the core problem of discovering how genes are distributed on the same chromosome remaining beyond the scope of that amended law.

(R2) Suppose that the reductionist could, after all, specify the content of classical genetics in terms of some set of fundamental laws. Deriving these from molecular biology would require the use of bridge laws, of the following sort:

$$(*) (x) (x \text{ is a gene} \leftrightarrow Mx),$$

where Mx is a (possibly complex) predicate of molecular genetics.²² Kitcher claims that molecular biologists do not offer any such claims as $(*)$ and that no such bridge laws can be found. We said earlier that the molecular analogue to classical (Mendelian) genes are segments of DNA. But which segments of DNA count as genes? Genes come in many lengths, and many DNA segments of those lengths are not genes; so "number of nucleotide pairs" will not suffice. Nor can we reckon a gene to be any segment of DNA between nucleotide triplets (codons) that initiate

and terminate transcriptions; mutations can produce single alleles containing within them terminating and reinitiating triplets, and moreover not every gene is transcribed onto the RNA figuring in protein synthesis. In short, no general structural statement can designate which DNA segments count as genes.

Despite the impossibility of defining the term *gene* in molecular terms, the reductionist might insist that, since there are only finitely many organisms and hence finitely many genes, it should be possible in principle to supply a list containing molecular descriptions of all such DNA segments. We might then construe Mx as the large disjunction of all such molecular descriptions. Here, Kitcher argues that the reductionist will have failed to provide us with anything sufficient to reduce a classical genetic law about genes or gene transmission. Laws, presumably, apply not only to actual cases but to cases that might have but did not actually arise: laws, we say, support their counterfactuals. (For more on this feature of laws, see the commentary on chapter 7.) Thus, to reduce a law about genes is to show how possible but nonactual genes would satisfy it. But the disjunctive approach just described, expressed in terms of all actual gene-structures, fails to do this.

Now the reductionist might suggest that biconditional bridge laws such as (*) are stronger than we need. For if we are after the derivation of the law

$$(1) (x)(y) ([Gx \& Gy] \rightarrow Axy)$$

(an amended version of Mendel's second law, say, about the independent assortment of genes), then perhaps we can derive (1) from (2)—some statement expressing merely necessary molecular conditions for being a gene—together with another lawlike premise (3):

$$(2) (x) (Gx \rightarrow Mx)$$

$$(3) (x)(y) ([Mx \& My] \rightarrow Axy).$$

Molecular genetics can presumably supply weaker claims of the form (2) to the effect that x is a gene only if it is composed of either DNA or RNA.²³ But Kitcher argues that there is no more hope of providing a strong molecular-biological law governing the independent assortment of all DNA and RNA segments, as in (3), than there is of providing a law of form (*) itself.

(R3) Suppose that the reductionist can provide fundamental laws of classical genetics and suitable bridge principles for deriving them. Our reduction model requires that the derivation of these laws explains them. Consider again an amended version of Mendel's second law. Why do genes of nonhomologous chromosomes assort and transmit independently? Cytology explains it perfectly in terms of the behavior of chromosomes

during meiosis. Kitcher argues that, while we can provide molecular genetic details involved in the cytological account, even to the point of a successful derivation of the transmission law, doing so would not elaborate or further explain the transmission law itself. Indeed, the further details would only obscure a perfectly good cytological explanation with irrelevant information.

The reductionist might object here that, even if further molecular details only obscure the explanation, this obscurity is simply an artifact of our limited cognitive abilities; the explanatory force of the molecular details could be grasped by more powerful intellects. Kitcher's response is that this misses the main point of his argument—it is not that the molecular derivation is too complicated but that it forfeits something present in the cytological explanation. The cytological account treats meiosis as belonging to a natural kind, namely, the kind of process that separates paired items and assigns them to distinct entities. In the meiotic instance of this process, homologous chromosomes are separated and assigned to distinct gametes. Kitcher's point is that explaining the transmission law requires identifying this pair-separation (or PS-) process as a natural kind to which meiosis belongs and that, from a molecular point of view, the PS-process forms no natural kind. From the viewpoint of molecular mechanics, PS-processes are a mixed bag; there are no general constraints on the molecular structures of items that can be paired or on how forces combine to pair or separate them. Since PS-processes, from the molecular viewpoint, are realized in so many disparate ways, a molecular account must fail to capture that feature of an adequate explanation of gene transmission that marks it as a species of a general kind. Hence, the cytological explanation, which uniformly accounts for a wide range of cases (meiosis among them) that from the molecular viewpoint are heterogeneous, is the objectively better explanation.²⁴

Not all philosophers of science would agree with Kitcher's denial of the explanatory thesis (R3). C. Kenneth Waters, for example, has explicitly rejected what he calls the "gory details" objection to reductionism²⁵—a reference to Kitcher's claim that "the distribution of genes to gametes is to be explained, not by rehearsing the gory details of the reshuffling of the molecules, but through the observation that chromosomes are aligned in pairs just prior to meiotic division, and that one chromosome from each matched pair is transmitted to each gamete" (1993). What force is there to Kitcher's point that the very broad natural kind he calls PS-processes, of which meiotic segregation is an instance, is heterogeneous from the molecular perspective? In claiming that pair separation may occur because of any number of forces, including the action of electromagnetic, nuclear, and even gravitational forces, Kitcher is surely conceiving of a natural kind that includes processes quite unlike anything that occurs during meiosis. Waters's first point is that, whatever analogies might hold between meiosis and other pair-separation processes of quite different sorts, it does not yet

follow that cytological theory offers a unified explanation of a broad range of phenomena that from the molecular viewpoint appear heterogeneous.²⁶ Of course, phenomena do exist that classical genetics can explain uniformly, phenomena that involve a relatively wide range of molecular mechanisms. Under the category of dominance, for example, classical genetics includes genes coding for structural proteins and enzymes, as well as regulatory genes. Waters's second point is that even if the molecular mechanisms by which these different sorts of genes are expressed are quite different, it is unclear why we should judge the shallower explanations of classical genetics preferable to the deeper accounts provided by molecular genetics. Moreover, even granting that the classical explanation is more unified, Waters believes that Kitcher is unduly pessimistic about molecular genetics, with the help of auxiliary assumptions, eventually giving an equally good explanation. That molecular genetics, apart from such assumptions, reflects the real diversity in a wide range of mechanisms should not be held against it.

KITCHER'S DIAGNOSIS OF THE FAILURE OF REDUCTIONISM

Quite apart from the difficulty of explaining classical genetics in molecular terms, Kitcher believes that the fundamental reason for the failure of reductionism lies with the falsity of (R1): according to Kitcher, we lack an account of theories that can be suitably applied to classical and molecular genetics. For if we adopt the standard view that a theory is a set of statements about laws and particular facts, then it is difficult to decide which statements make up the theory of classical genetics. What we find in the classical genetic literature at any given time are not laws but claims about particular organisms—paradigmatic organisms such as fruit flies, bread molds, and maize. So either classical genetics is merely a motley collection of statements about various particular organisms, or we must suppose that although there are general laws in genetics, these laws are never enunciated by geneticists. Two questions thus arise: What exactly is the relationship between the collection of statements about particular organisms accepted at a given time and the theory of classical genetics at that time? And what is the relation among versions of the theory at different times that makes us reckon them versions of the same theory?

Adopting the approach of T. S. Kuhn to understanding science through paradigms, Kitcher rejects the standard view that a theory is nothing more than a set of statements. The set of statements about (say) inheritance at some time is only one component of a far more complicated entity Kitcher calls the *practice* of classical genetics at that time. The practice of classical genetics is determined by a common language about heredity, a set of statements or beliefs about inheritance held at a time, a set of questions taken to be appropriate about heredity, and abstract patterns of reasoning instantiated by attempts to answer those questions. The

beliefs exemplify the theory of classical genetics insofar as they figure in particular problem-solutions arising from current practice; by applying certain patterns of reasoning, one puts forward answers to accepted questions. Versions of the theory we call *classical genetics* are versions of the same theory because different practices are linked by relatively small changes in language, accepted questions, and patterns of reasoning. Moreover, these versions of the theory share a common task, namely, the goal of characterizing hereditary phenomena. Thus, teaching students classical genetics is not a matter of giving them some set of fundamental axioms or laws; rather, it involves explaining a vocabulary, posing a series of crucial questions, and training students to answer those questions by using certain patterns of reasoning.

Kitcher illustrates this view of classical genetics by describing the initial family of questions to which the theory is applied. This is the family of pedigree problems that arise when we observe various phenotypic distributions among several generations of organisms. Each such problem is characterized by a set of data, a set of constraints, and a question. The data are statements about the distribution of phenotypes among the organisms of the pedigree. The constraints are general cytological facts and descriptions of the chromosomal constitution of members of the species. The questions are about what distribution of phenotypes we should expect from a cross between two individuals of the pedigree, about the probabilities that some phenotypic characteristic will arise from such a cross, and so on. With such problems in hand, one advances to a solution by offering a genetic hypothesis specifying what alleles are relevant, how they are phenotypically expressed, and how they are transmitted through the pedigree. One then answers the questions about the distribution and probability of phenotypes by using the hypothesis to compute the expected distribution of genotypes among offspring.

According to Kitcher, the history of classical genetics is the history of versions of this kind of practice, applied to pedigree and other families of problems. Differences among successive versions of classical genetic theory arise from differences in their genetic hypotheses and the new constraints placed upon them.

Returning now to the relation between classical genetics and molecular genetics, Kitcher is concerned to address how the explanations of hereditary phenomena offered by molecular biology connect with those offered by classical genetics. He describes three major contributions of molecular genetics to our understanding of hereditary phenomena. The first is an account of gene replication, to which classical genetics was committed without any accompanying explanation. As we saw earlier in our brief review, the Watson-Crick model of DNA explains how replication is possible. A second major advance was a molecular-genetic account of mutation. By understanding genes as segments of DNA, we can recognize possible mechanisms whereby mutant alleles are produced by the

addition, deletion, or substitution of nucleotides during replication, leading to deviant proteins and, ultimately, to deviant phenotypes. Finally, Kitcher cites the specific explanation of how a substitution error in transcription forms the mutant allele responsible for the deviant hemoglobin in cases of sickle-cell anemia. The resulting change in one of the four amino-acid chains making up the hemoglobin molecule alters its ability to bind with and give up oxygen. These three examples, understood in light of Kitcher's view of theories described above, reflect three different sorts of relations among successive theories: presupposition, conceptual refinement, and explanatory extension. Each of these relations departs from the standard account of intertheoretic reduction in significant ways. Let us consider them in turn.

PRESUPPOSITIONS: THE EXAMPLE OF GENE REPLICATION

Kitcher insists that the assertion that genes replicate is not a central law in classical genetics and compares it unfavorably to bona fide laws such as the Boyle-Charles law used in thermodynamics. His reasons for denying lawlike status to gene replication depend on his adopting a functional characterization of laws, since he indicates in a footnote that he is prepared to concede that "genes can replicate" is a true universal generalization. His point is that, even though it is a true universal generalization, "genes can replicate" is far too weak to explain anything interesting. Thus, the generalization fails to qualify as a law because it cannot perform one of the distinctive functions of laws within science, that is, to explain things in conjunction with other laws. (For a discussion of some of the problems with adding functional characteristics to the simple regularity theory of laws, see "Dretske's Criticisms of Ayer's Theory" in the commentary on chapter 7.)

According to Kitcher, gene replication does not appear as part of any classical-genetic explanation; rather, it is presupposed by such explanations. Before the advent of molecular genetics the presupposition of gene replication was problematic because it was hard to understand how such complicated molecules could duplicate themselves so exactly, especially when they had suffered damaging mutations. In answering the question, How is gene replication possible? molecular genetics thus gave a theoretical demonstration of a formerly problematic presupposition of classical genetics.

Kitcher's analysis of this process runs as follows. Theory *T* is said to presuppose *p* if instantiations of the problem-solving pattern of *T* contain statements implying *p*. Now, suppose that *T*-theorists recognize an argument from accepted premises to the conclusion that *p* is impossible: in such a case, *p* is a problematic presupposition. If a new theory *T'* were to produce an argument showing that *p* is possible and showing what is wrong with the existing reasoning that threatens *p*, then we can say that

T' gives a theoretical derivation of the possibility of an antecedently problematic presupposition of T . Kitcher argues that molecular biology provides just such a derivation of the possibility of gene replication. But he notes that this account of the relation between T and T' differs from the standard concept of reduction in two ways. First, it requires no commitment to the view that T and T' can be formulated as sets of fundamental laws (and their consequences). Second, it requires no commitment to the notion that all general claims about genes in T are equally in need of derivation from T' ; rather, the derivation is an isolated case in which T' addresses some particular thesis that is problematic but that underlies all T -theoretic explanations. Thus, where the reductionist looks for a global relation between T and all fundamental laws of T' , Kitcher's account of the relation between classical and molecular genetics focuses on particular derivations of problematic presuppositions in classical genetic theory.

CONCEPTUAL REFINEMENT: THE EXAMPLE OF MUTATION

The referent of the term *mutant allele* in classical genetics was originally fixed through the description "chromosomal segment producing a heritable deviant phenotype." This description needed amending because it covered, not only the targeted internal changes in the structure of genes, but also cases of gene transposition in unequal crossing-over at recombination. Molecular genetics gives a precise account of the relevant internal changes in mutation (insertion, deletion, substitution), and in thus refixing the referent of *mutant allele* makes it possible to distinguish cases of mutation from the untargeted cases of recombination. In this way, the molecular-genetic account of mutation involves an intertheoretic relation that Kitcher calls the *conceptual refinement* of some theory T by a successor theory T' . Conceptual refinement is accomplished when T' specifies the entities within the extensions of predicates of T in a way that changes our means of fixing the referents of these predicates.²⁷ After the conceptual refinement wrought by molecular genetics, the term *mutant allele* no longer includes the cases of recombination that had confused earlier researchers.

EXPLANATORY EXTENSION: THE EXAMPLE OF SICKLE-CELL ANEMIA

The relation between classical and molecular genetics, as illustrated by the sickle-cell anemia example, might appear to approximate standard reductionist models. Statements ascribing particular phenotypes (ultimately, blockages of blood flow caused by sickle-shaped red blood cells packed with deviant hemoglobin) to particular genotypes (the mutant allele resulting from a nucleotide substitution) would, it seems, be derivable from the molecular characterization of the relevant allele. Encouraged by this

apparent success of the standard reductionist model, we might suppose that some general pattern of reasoning in molecular genetics (about the way in which alleles produce proteins) would enable us to generate conclusions about the classical association between each genotype and its corresponding phenotype.

Kitcher proposes that we understand the relation between the patterns of reasoning found in different theories by means of the notion of an explanatory extension of one theory by another. A theory T' is said to provide an explanatory extension of T only if one of the premises of some problem-solving pattern of T can be generated as the conclusion of some new problem-solving pattern of T' . Relevant premises occurring in explanations by T would themselves be explained by explanatory extension arguments of T' . But, Kitcher argues, we cannot infer from this, as the reductionist would have it, that the old T -explanations can always be improved by replacing its relevant premises by the T' -derivations. For—in a way similar to that encountered in the failure of (R3)—what is relevant for the classical genetic explanations of some claim S may not be relevant for explaining some claim S' appearing in the explanatory derivation of S .

Despite its similarity to the derivation condition in the standard reductionist model, Kitcher denies that his relation of explanatory extension is simply reduction under another guise. True, it is a global relation among theories (unlike the replication case above), and it is formulated in a way that frees it from the requirement to accommodate (R1) and (R2). Indeed, in the hemoglobin case it appears that we can, with the assistance of certain limiting boundary conditions (as in the derivation of Galileo's and Kepler's laws from Newton's theory), derive the classical genetic statements about normal and mutant hemoglobin genes from molecular accounts of the genetic structure. But the simplicity of the sickle-cell anemia case makes it a highly atypical example of explanatory extension. Red blood cells are the simplest cells in the human body: lacking a nucleus, they are just membranes containing hemoglobin. The abnormal behavior of sickle cells is a direct consequence of their unusual shape, and their unusual shape is a direct consequence of the deviant hemoglobin molecules they contain. Similarly, the deviant hemoglobin molecules are the result of a single, point mutation on a single allele that has the production of the deviant hemoglobin as its sole effect. Normally, mutations operate in a far more complicated manner, affecting many different phenotypic characteristics, such that derivations of genotype-phenotype associations from molecular genetics alone are not possible. Any appearance of reductionistic promise from explanatory extension is an artifact of the unusual simplicity of the hemoglobin case.

8.5 | Summary

Philosophers of science have made many attempts to characterize the relation between earlier and later scientific theories by using the concept of reduction. Ernest Nagel's influential account characterizes the reduction of one theory T by another T' as a species of explanation. Given that theories can be explicitly formulated as sets of meaningful statements, and given a deductive model of explanation, T reduces to T' on this account if the fundamental laws of T can be derived from those of T' . Two important features of reduction are entailed by this basic formulation: first, the claims of T and T' must be logically consistent, and second, terms shared by the vocabularies of T and T' must have a common meaning. These tenets, fundamental to Nagel's correspondence analysis of reduction, are called the consistency condition and the condition of meaning invariance.

The laws of a reduced theory often contain terms that are not shared by the reducing theory. This leads Nagel to distinguish homogeneous reductions from inhomogeneous (or heterogeneous) reductions. In inhomogeneous (as opposed to homogeneous) reductions, the descriptive terms of the reduced theory T are not a proper subset of the vocabulary of the reducing theory T' . Nagel makes clear that in such inhomogeneous reductions the explanatory derivation can be effected only by supplementing the reducing premises of T' with some additional statements connecting the unique vocabulary of T with the terms of T' . These additional premises are called *bridge laws*. Since the distinctive terms of a reduced theory T have meanings independent of their relation to the new reducing theory T' , bridge laws cannot be construed as expressing any analytic or semantic connection between the two vocabularies. Instead, Nagel argues, bridge laws are empirical hypotheses that may take one of two forms. Either the bridge laws state—in terms of T' —conditions under which the properties expressed by the distinctive terms of T are manifest or they assert that apparently distinct objects or processes are in fact identical.

Paul Feyerabend rejects Nagel's correspondence analysis and, along with it, much else of what Feyerabend regards as misguided and pernicious in the logical empiricist account of scientific methodology. Two cornerstones of this methodology are the consistency condition and the condition of meaning invariance, both of which are crucial to Nagel's model of intertheoretic reduction. According to the consistency condition, some proposed theory about a domain of phenomena is to be judged acceptable only if it is consistent with theories already accepted in that domain. According to the condition of meaning invariance, meanings of scientific terms must not alter as science progresses and theories change. Feyerabend argues that these conditions are unreasonable constraints on science that are often violated by actual scientific practice.

Why are these conditions unreasonable? Feyerabend argues that the consistency condition militates against the acceptance of some alternative theory T^* on quite arbitrary grounds—on its temporal relation to T , which just so happens to be the earlier of the two, not its relation to the facts—and so it promotes an unreasonable methodology that no genuine empiricist should tolerate. If we try to judge whether T^* represents an advance over T on the basis of “the facts,” we are guilty of assuming the relative autonomy of facts—that the facts exist and are available to us regardless of whether we have already considered alternatives to T . Feyerabend insists that all facts are theory dependent. Thus, the invention of alternative theories in disagreement with an existing theory T —which the consistency condition forbids—is crucial to a properly empiricist methodology. Finally, Feyerabend argues, application of the consistency condition contributes to a false sense of the “success” of existing theories, perpetuating T by a kind of methodological conformism. In the case of the meaning-invariance condition, Feyerabend claims that a methodology allowing mutually inconsistent (but empirically adequate) theories would welcome the use of mutually irreducible theories whose terms are unconnected by anything like factually correct bridge laws. Welcoming such theories brings with it no demand that meanings remain invariant.

Not only are the consistency and meaning-invariance conditions unreasonable restrictions on scientific change, but they are also, according to Feyerabend, violated by actual scientific practice. Against the consistency condition, Feyerabend argues that the relevant (terrestrial and celestial) consequences of Newtonian theory are inconsistent with Galileo's and Kepler's laws: strictly speaking, the Newtonian theory is a replacement for, not a reduction of, the laws of Galileo and Kepler. Against the condition of meaning invariance, Feyerabend argues that, in claiming that the mass of an object depends on its velocity, relativity theory means by *mass* something different than what classical mechanics means by this term. (According to classical mechanics, mass is independent of velocity.) As before, the old theory is not reduced to the new theory in Nagel's sense but is replaced by it. This result applies not only to theoretical terms, but to observational terms as well since observational terms, no less than theoretical terms, are theory laden.

Nagel rejects Feyerabend's claim that the consistency and meaning-invariance conditions are falsified by actual scientific practice. There is no fundamental threat to consistency in granting that reducing theories entail close approximations to the laws of reduced theories, since new theories often specify special conditions under which those approximations hold. Moreover, in actual practice derivations of laws commonly invoke simplifications and approximations of various sorts. Regarding Feyerabend's denial of meaning invariance, Nagel argues that it renders unintelligible any claims to the effect that T and T' are consistent or inconsistent—these being logical relations holding between statements

whose terms have shared meanings. Furthermore, in the absence of shared meanings, we cannot intelligibly express the fact that the domains of phenomena to which T and T' apply partially overlap: if any description of the domains is theory laden, there are no means by which to express the claim that T' covers all the phenomena covered by T .

Our discussion of reduction thus far has supposed that there is a unified class of reductive phenomena and that our task is to formulate an account that captures its basic features. Thomas Nickles argues that the relations among theories are far too diverse to suppose that paradigm examples of reduction form a unified class for which a single model is available. Many standard examples are indeed reductions between a narrower, less fundamental theory T and some broader, more fundamental theory T' . In such cases, T is consolidated by T' to effect a conceptually and ontologically more efficient system. Such a relation, between two mutually consistent and often concurrent theories, Nickles calls reduction₁. But not all reductions are of this domain-combining sort. In a large class of standard examples, it is far more natural to see the broader, more general theory T' as reducing, under some limiting transformation, to the narrower, more restricted theory T . Special relativity reduces in this sense to classical mechanics, in the limit of low velocities. Such domain-preserving reductions, between logically incompatible earlier and later theories, Nickles calls reductions₂.

Nickles's distinction between two kinds of reduction provides a useful way of thinking about Nagel's account of intertheoretic reduction. Reductions₁ will generally be inhomogeneous and, using bridge laws, will involve the derivational explanation of T by T' ; reductions₂ will typically be homogeneous and never derivational. Thus, Nagel's account captures only reduction₁ and, in treating all reduction as derivational, casts the phenomenon of intertheoretic reduction in too narrow a mold. In doing so, Nagel opens his account to Feyerabend's objections. While Nickles does not explicitly criticize Feyerabend's rejection of meaning invariance, he does imply that Feyerabend's doctrine of meaning change applies at most to homogeneous reductions₂, not to shared terms in the case of domain-combining inhomogeneous reductions₁; the meaning of *mass* may change from classical to relativistic physics, but *pressure* and *volume* retain a univocal sense in both thermodynamics and statistical mechanics. Likewise, an across-the-board denial of the consistency condition is too strong. In some cases, incompatibilities among theories are historical accidents, not conceptual requirements of the reduction. But in allowing a class of reductions₂ that, in departing from Nagel's derivational model, do not require the consistency condition, Nickles's account is immune to Feyerabend's consistency objection, which asserts that reduction is impossible when a successor is inconsistent with its predecessor.

If the existence of reductions₂ show that Nagel's conditions are not necessary for intertheoretic reduction, other examples show that they are

not sufficient. In many cases, biconditional statements connecting two theories would, on Nagel's derivational account, trivially allow a reduction where it clearly does not exist—as when the inclusion of mechanical laws in the molecular-kinetic theory of matter would allow for the derivation of mechanics from the kinetic theory (and hence the putative “reduction” of mechanics to the kinetic theory). In other cases, in which one might view bridge laws as specifying that entities previously thought to be distinct now be identified, Nickles suggests that heuristic and metatheoretical considerations count in favor of a much weaker relation between two theories. It is precisely such cases that the notion of reduction, is meant to capture, since these are examples in which conceptual and ontological economy is not at issue.

Most of the standard examples of intertheoretic reduction have been drawn from the physical sciences. To what extent are recent developments in biology instances of reduction? In particular, is the relation between classical genetics and molecular biology sufficiently similar to standard examples in the physical sciences to justify the claim that classical genetics has been reduced to molecular biology?

Philip Kitcher denies any reductionist account of the relation between these two biological theories. Given a familiar model of reduction as a derivational explanation by T' of the fundamental laws of T with the help of suitable bridge laws, the reductionist position would require acceptance of three theses:

- R1** Classical genetics contains laws about gene transmission that are conclusions of reductive derivations from molecular genetics.
- R2** The vocabulary of classical genetics can be linked by suitable bridge laws to the vocabulary of molecular genetics.
- R3** A derivation of laws about gene transmission from principles of molecular biology explains why these laws hold (to the extent that they do).

Kitcher rejects all three of these theses, arguing that there are no laws in classical genetics, that no bridge laws connect terms in classical genetics with terms in molecular genetics, and that no laws (or even approximate laws) of classical genetics can be derived from molecular biology.

Concerning (R1), Kitcher claims that general laws are largely absent from the classical genetic literature. Where they do occur, they play no crucial role in the actual practice and articulation of classical genetics.

Regarding (R2), putative bridge laws of the form “ x is a gene if and only if Mx ” (where M is a predicate from molecular genetics) are never formulated by molecular biologists, and for good reason: no adequate predicates from molecular genetics are to be found. In particular, no general structural statement can specify which segments of DNA count as genes.

In connection with (R3), Kitcher argues that while molecular biologists can provide the “gory details” of how gene transmission occurs at the molecular level, these details do not explain anything that we did not already understand perfectly well using the general principles of classical genetics and cell biology. For example, our explanatory understanding of why genes segregate at meiosis depends on the process of chromosome separation and migration. This is an instance of what Kitcher calls the PS- (or pair-separation) *process*. The explanatory power of pair separation depends on its being a general kind of process that occurs at the level of cells, not at the level of molecules. The many different things going on at the molecular level during particular instances of pair separation are irrelevant to the unified explanation that it gives of gene segregation at the cellular level. One of the fundamental errors of reductionism is to think that, just because systems are composed of smaller components, an explanation at the level of the components must be better than an explanation at a higher level. Often, especially in biology, the reverse is true.

Kitcher traces the failure of the standard model of reduction in genetics to our lack of an adequate account of scientific theories and the relations among them. He proposes that we think of classical genetics as a practice rather than as a set of statements. The practice of classical genetics at a given time includes some common language about heredity, a set of beliefs about inheritance, a set of appropriate questions about heredity, and patterns of reasoning that scientists use to answer those questions. Teaching students classical genetics is not teaching them laws or axioms but training them in this practice. Typically, research involves problems made up of a set of data, a set of constraints, and a question. To solve those problems, a genetic hypothesis will be proposed and then tested against its predictions.

What, then, is the relation of molecular genetics to the practice of classical genetics as described by Kitcher? From three important contributions of molecular to classical genetics, Kitcher isolates three sorts of relations that can hold between the theories: presupposition, conceptual refinement, and explanatory extension.

First, classical genetics will sometimes be committed to, but be unable to account for the possibility of, some general proposition p . A new theory (such as molecular genetics) can sometimes provide the premises from which the possibility of that presupposition can be derived. Unlike the reductionist account, this requires neither that p be a law of classical genetics, nor that all general claims of the classical theory be equally in need of derivation.

Second, a new theory like molecular genetics will sometimes provide a new means (in its vocabulary) of fixing the referent of terms in the older theory. Kitcher calls this relation *conceptual refinement*.

Third, a new theory such as molecular genetics can sometimes provide an explanatory extension of the older classical genetics, by using some

pattern of reasoning to derive general statements of the older theory. But from such an explanatory extension we cannot infer, as the reductionist requires, that classical-genetic explanations can be improved by replacing their premises with the new derivations offered by molecular genetics because what is relevant to a classical-genetic explanation of some claim *S* may not be relevant to explaining some claim *S'* figuring in the explanatory derivation of *S*. In short, a range of intertheoretic relations hold between classical and molecular genetics, none of which are, or contribute to, an instance of reduction as standardly conceived.

■ | Notes

1. The classic account of intertheoretic reduction is offered in chapter 11, "The Reduction of Theories," of Ernest Nagel's *The Structure of Science* (New York: Harcourt, Brace and World, 1961). This chapter, in which Nagel is most explicit about the formal conditions of reduction, is a revised version of his earlier "The Meaning of Reduction in the Natural Sciences," *Science and Civilization*, ed. R. C. Stauffer, (Madison: University of Wisconsin Press, 1949).
2. Any argument in which the same word appears in two or more premises, or in at least one premise and the conclusion, but in different senses, is said to commit the fallacy of equivocation. Take the following argument, for example: "Sugar is a vital ingredient in all human cells. Cotton candy contains sugar. Therefore, cotton candy contains something which is a vital ingredient in all human cells." The equivocation on the word *sugar* renders the argument invalid. In the first premise it means *glucose*; in the second premise it means *sucrose*. Strictly speaking, in defining the fallacy of equivocation we should also stipulate that the word in question be essential to the derivation of the conclusion: merely adding an ambiguous word to the premises and conclusion of an argument that is already valid will not produce an invalid argument.
3. The exceptions: from "*P*" we can validly infer "*P* or *Q*," and from "All dogs are mammals" we can validly infer "All brown dogs are mammals". Nagel explains the irrelevance of such exceptions in chapter 11 of *The Structure of Science* (1961).
4. For more on the role of observation terms in reduction, see the section "The Apparent Violation of the Condition of Meaning Invariance" later in this commentary. The theory-observation distinction is explored at length in chapter 9.
5. Bridge laws are sometimes called *bridge principles*, *correspondence rules*, or *co-ordinative definitions*.
6. If the stated *T'* conditions are only necessary (rather than necessary and sufficient) for the occurrence of the relevant property, then the laws of *T* in which the property figures could not be derived from the claims of *T'* conjoined with bridge laws. But Nagel's third assumption of the correspondence analysis of inhomogeneous reduction is that reductions are deductive explanations of the reduced laws.

7. In his paper, Feyerabend uses *T* to stand for the reducing (explaining) theory, and *T'* for the theory that is reduced (explained). This is exactly the opposite of Nagel's notation: for Nagel it is *T'* that reduces (explains) *T*. To avoid confusion, we have adopted Nagel's notation throughout this commentary, except when quoting Feyerabend on pp. 1017–18.

8. In forming a proper assessment of Feyerabend's paper, it is important to remember that when it was published (in 1963), logical empiricism was still a dominant force in philosophy of science. Kuhn's book, *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1962) had only just appeared, and many of Kuhn's arguments (especially those about observation and meaning) are substantially indebted to Feyerabend's work. Historically, Feyerabend's criticisms of logical empiricism played a major role in its demise. The chorus of disapproval that has greeted Feyerabend's later writings—books such as *Against Method* (London: New Left Books, 1975) and *Science in a Free Society* (London: New Left Books, 1978)—because of their gleeful embrace of relativism and their anarchistic rejection of the authority of science (“anything goes”), should not obscure the significance of his earlier work as a critic of logical empiricism.

9. For a detailed account of attempts to explain Brownian motion in the nineteenth century, see Mary Jo Nye, *Molecular Reality* (London: Macdonald, 1972).

10. The simplest of these predictions was that the kinetic energy of the pollen grains should equal the kinetic energy of the water molecules surrounding them. But the pollen grains were found to be moving much more slowly than predicted. Eventually, Einstein was able to clear up the mystery by pointing out that when we follow the zig-zag path of a pollen grain under the microscope, what we observe is never the instantaneous velocity of the pollen grain (which decays in less than a millionth of a second) but a much smaller, average velocity.

11. For a criticism of Feyerabend's interpretation of this history, see Ronald Laymon, “Feyerabend, Brownian Motion, and the Hiddenness of Refuting Facts,” *Philosophy of Science* 44 (1978): 225–47.

12. The second law of classical thermodynamics says that the entropy of an isolated system never decreases. It claims to be a true, universal law of nature. As Maxwell and Boltzmann realized, the kinetic theory implies that there should be random, local fluctuations in entropy even in a fluid that, as a whole, is in thermal equilibrium. Most of these spontaneous variations in density and temperature will be very small and last for only a fraction of a second; the larger the variation, the more improbable its occurrence. Thus, from the standpoint of the kinetic theory, the second law of thermodynamics is not a universal, lawlike truth but a statistical one. Interestingly, although Maxwell and Boltzmann were the two foremost advocates of the kinetic theory in the nineteenth century and the pioneers of statistical mechanics, neither of them thought that the tiny random fluctuations predicted by the theory would ever be observed. Apparently, the significance of Brownian motion never occurred to them.

13. The relevant facts included numerical predictions of how far the particles would move in a given time, how fast they would rotate, and how the number of particles per unit volume of fluid in a vertical column would decrease with height.

14. See the final paragraphs of the section "The Apparent Violation of the Condition of Meaning Invariance," later in this commentary, for a discussion of Nagel's response to this criticism of synonymy and meaning invariance.

15. Kepler's third law asserts that for all the planets, the square of the period divided by the cube of the mean distance from the sun is a constant. Newton's theory entails that this ratio is not constant because it depends on the planet's mass.

16. The mass of an object, according to special relativity, is given by $m = m_0 + \frac{1}{2}m_0v^2/c^2$, where m_0 is its rest mass and c is the speed of light. For velocities much lower than c , $m \approx m_0$ within the limits of experimental error, but for velocities nearer c , the difference between m and m_0 becomes appreciable.

17. From p. 59 of "Explanation, Reduction, and Empiricism," in *Scientific Explanation, Space, and Time*, ed., H. Feigl and G. Maxwell, vol. 3, *Minnesota Studies in the Philosophy of Science*, (Minneapolis: University of Minnesota Press, 1962), 28–97.

18. Length contraction is the well-confirmed phenomenon, predicted by Einstein's special theory of relativity, that moving rods shrink in the direction of their motion when their length is measured by an observer who is not traveling with them. For example, suppose that I have two rods, each exactly one meter long as they lie on the ground at my feet. If one of the rods is now set in motion and I measure its length as it flies past me, it will be less than one meter. Moving rods contract by a factor of $1/\sqrt{1 - v^2/c^2}$, where v is the velocity of the rod relative to me, the observer, and c is the velocity of light. A person travelling with the rod will make the same judgment about the rod lying at my feet. According to that person, my rod has shrunk by the same fraction of its length. The effect becomes noticeable only when the velocity of the rod, v , approaches the velocity of light, c .

19. Lawrence Sklar, "Types of Intertheoretic Reduction," *British Journal for the Philosophy of Science* 18 (1967): 109–24.

20. One operation is the inverse of another if it completely reverses the effect of the other when both operations are applied in turn. Since a limiting operation generates classical mechanics when applied to the special theory of relativity, its inverse would have to generate the special theory of relativity when applied to classical mechanics. Einstein's theory is transformed into Newton's by letting the speed of light become infinite. The relativistic Lorentz factor $1/\sqrt{1 - v^2/c^2}$ in Einstein's theory becomes equal to 1 and so drops out of the equations. Nickles's point is that there is no well-defined operation that would enable us to recover the original theory from the transformed equations because the original theory contains factors and constants that are absent from those equations. The limiting operation involved in this kind of reductive transformation does not have an inverse.

21. Hereafter in this section we will use the terms *reduction* and *reductionism* as shorthand for "reduction of classical to molecular genetics" and for "classical genetics reduces to molecular genetics," respectively, dropping the explicit reference to genetics.

22. Kitcher uses the symbol " \leftrightarrow " to stand for the connective "entails and is entailed by." Thus, (*) is to be read as "x is a gene" entails that x has property M, and "x has property M" entails that x is a gene.
23. Viral genes are composed of RNA.
24. Kitcher is here relying on a view of theoretical explanation as unification that he has defended elsewhere. See Philip Kitcher, "Explanatory Unification and the Causal Structure of the World," in *Scientific Explanation*, ed. P. Kitcher and W. C. Salmon, vol. 13, *Minnesota Studies in the Philosophy of Science* (Minneapolis: University of Minnesota Press, 1989), 410–505, and Philip Kitcher, "Darwin's Achievement," chap. 2, *The Advancement of Science* (Oxford: Oxford University Press, 1993).
25. C. Kenneth Waters, "Why the Antireductionist Consensus Won't Survive the Case of Classical Mendelian Genetics," in *PSA 1990*, vol. 1, ed. A. Fine, M. Forbes, and L. Wessels (East Lansing, Mich.: Philosophy of Science Association, 1990), 125–39.
26. One might add here that for any wide range of phenomena there is likely to be some abstract description under which they all fall and that such a "natural kind" will undoubtedly appear heterogeneous from many perspectives. This articulation of a putative natural kind would be vindicated to the extent that some general theory encompassing all such instances uses the natural kind in its explanations. Waters's first point is that cytological theory accomplishes nothing of that sort.
27. Kitcher does not discuss the relation of conceptual refinement to the standard model of reduction. Could we view statistical mechanics as offering a conceptual refinement of the term *temperature* in thermodynamics? Presumably not. Kitcher's account of conceptual refinement seems to presuppose that the predicate whose means of reference-fixing is changed in going from *T* to *T'* is present in the vocabulary of both *T* and *T'*. While this would appear to be so in the case of *mutant allele*, it is not so in the case of *temperature*.



9

Empiricism and Scientific Realism

INTRODUCTION

One of the most lively and persistent debates in twentieth-century philosophy of science is that between empiricists and realists concerning the status of scientific theories. Are scientific theories to be understood as offering (or at least intending to offer) a true account of the world? And if we accept scientific theories as true, should we believe that the entities they appear to postulate really exist?

Empiricists emphasize that the warrant for all scientific claims ultimately rests on experience, on what can be observed, tested, and measured. They are thus suspicious of theoretical assertions that cannot in some fairly direct way be cashed out in terms of experience. This suspicion has, in the hands of some, been reinforced by arguments from underdetermination: given the same body of observational evidence, there are in principle an indefinitely large number of alternative theories that are compatible with that evidence and which entail it. This being so, what could warrant our selecting just one of those theories and asserting that it is true or even probable? Moreover, reflecting on the frequency with which past scientific theories have later been proved false would seem to recommend a cautious skepticism or agnosticism about the unobservable entities postulated by present theories.

Realists concede that many theories accepted in the past have turned out to be false. Nevertheless, they insist that there is a clear sense in which science progresses and improves over time. Present-day theories in physics, chemistry, and biology are objectively better than theories of the eighteenth and nineteenth centuries. Even if present-day theories are not true in every detail, they are closer to the truth, and this makes them more reliable for making predictions and guiding our actions. Although theories are, of course, tested by experience, their claims are not merely about

experience, nor are they in some way reducible to it. The fundamental point of theorizing is to go beyond experience to discover deep, hidden truths about the underlying causes of events, regardless of whether these causes can be directly observed. And when such theories are well confirmed by the observable evidence they predict, we have the best reason possible for taking their theoretical claims at face value and for believing that the entities they postulate are real. In this way science has extended our knowledge into the smallest parts of atoms and the farthest reaches of the universe.

Many empiricist philosophers of science are antirealists about the purely theoretical claims of science and the unobservable entities (such as electrons and quarks) to which its theoretical terms allegedly refer. In "The Ontological Status of Theoretical Entities," Grover Maxwell argues that this antirealist view of science cannot be sustained. In particular, Maxwell argues that what counts as observational is itself a theoretical question and hence that the traditional observational-theoretical distinction at work in empiricist accounts of science has no ontological significance whatsoever: no entities are unobservable in principle.

In "Arguments Concerning Scientific Realism"—a piece taken from his 1980 book *The Scientific Image*—Bas van Fraassen argues against Maxwell's position on the way to formulating his own constructive empiricism. According to van Fraassen's version of antirealism, we can accept scientific theories and yet remain agnostic about their truth, requiring only that theories be empirically adequate (i.e., true about observables). From the standpoint of constructive empiricism, van Fraassen offers his diagnosis of why a number of important arguments for realism must be regarded as failures.

Alan Musgrave criticizes van Fraassen's approach to scientific theories in his "Realism versus Constructive Empiricism." One of the most popular arguments for scientific realism, which, according to Musgrave, can survive van Fraassen's attack, is the "miracle" or "ultimate" argument—that if we do not accept our scientific theories as (at least approximately) true and their terms as in fact referring to unobservables, then we can only reckon the success of our theories to be a miracle. In "A Confutation of Convergent Realism," Larry Laudan attacks this inference-to-the-best-explanation defense of realism, arguing from the history of science that the connection between approximate truth and referential success is much weaker than realists suppose. Laudan objects that the notion of approximate truth, or verisimilitude, is too imprecise to bear the weight that realists place on it. Laudan's objection is discussed at length by James Robert Brown in "Explaining the Success of Science." Brown agrees that we do not have any analysis of verisimilitude that is able to sustain the traditional success-of-science argument for realism. In its place, Brown proposes a weaker kind of realist explanation of the success of science, one in which that success does not also justify our belief in realism.

How then might one try to justify scientific realism? One might narrow one's sights, dropping the troublesome notion that theories are true or approximately true. In "Experimentation and Scientific Realism," Ian Hacking urges that we sever the view that theories are true from a realist commitment to the existence of unobservables. According to Hacking's experimental realism, our justification for believing that unobservables exist rests on our ability to manipulate them in scientific experiments. David Resnik replies to this proposal in "Hacking's Experimental Realism," where he argues that Hacking's entity realism can be defended only by relying on some form of traditional success-of-science argument. In addition, Resnik argues that the only initially plausible way of trying to wed realism about entities to antirealism about theories requires a theory of meaning that brings realism about theories along with it after all.

Arthur Fine suggests that we hunt for some middle ground between scientific realism and antirealism. After criticizing arguments for realism in "The Natural Ontological Attitude," he recommends a broadly minimalist attitude toward scientific theories, the natural ontological attitude (NOA), according to which the confirmed results of science are accepted in just the same way we naturally accept the evidence of our senses about garden-variety objects. According to Fine, this is a nonrealist proposal, a core position to which realists and antirealists add unnecessary and indefensible philosophical accounts of truth. In "NOA's Ark—Fine for Realism," Alan Musgrave responds to Fine by arguing that on any plausible construal of NOA, it is not a nonrealist philosophy of science at all but a thoroughly realist one. Musgrave then offers an alternative construal of Fine's account that is genuinely neither realist nor antirealist, but, Musgrave argues, this alternative account is devoid of philosophical content.

The Ontological Status of Theoretical Entities

That anyone today should seriously contend that the entities referred to by scientific theories are only convenient fictions, or that talk about such entities is translatable without remainder into talk about sense contents or everyday physical objects, or that such talk should be regarded as belonging to a mere calculating device and, thus, without cognitive content—such contentions strike me as so incongruous with the scientific and rational attitude and practice that I feel this paper *should* turn out to be a demolition of straw men. But the instrumentalist views of outstanding physicists such as Bohr and Heisenberg are too well known to be cited,* and in a recent book of great competence, Professor Ernest Nagel concludes that “the opposition between [the realist and the instrumentalist] views [of theories] is a conflict over preferred modes of speech” and “the question as to which of them is the ‘correct position’ has only terminological interest.”¹ The phoenix, it seems, will not be laid to rest.

The literature on the subject is, of course, voluminous, and a comprehensive treatment of the problem is far beyond the scope of one essay.

FROM Herbert Feigl and Grover Maxwell, eds., *Scientific Explanation, Space, and Time*, vol. 3, *Minnesota Studies in the Philosophy of Science* (Minneapolis: University of Minnesota Press, 1962), 3–15.

* Niels Bohr and Werner Heisenberg, along with Max Born and other pioneers of quantum physics in the 1920s, were proponents of the so-called Copenhagen interpretation of quantum mechanics, which remains the orthodox textbook account to this day. Its elements include Bohr's doctrine of complementarity, the insistence that the wave function contains a complete description of reality, and a profound reluctance to say anything about the properties of quantum systems when they are not being observed or measured. As Maxwell says, Bohr and Heisenberg were instrumentalists (though opinions are sharply divided over what, exactly, Bohr's position was), and their views had a considerable influence on the development of logical positivism and logical empiricism. For a good account of Heisenberg's version of the Copenhagen interpretation, see Werner Heisenberg, *Physics and Philosophy* (New York: Harper and Row, 1958).

I shall limit myself to a small number of constructive arguments (for a radically realistic interpretation of theories) and to a critical examination of some of the more crucial assumptions (sometimes tacit, sometimes explicit) that seem to have generated most of the problems in this area.²

■ | The Problem

Although this essay is not comprehensive, it aspires to be fairly self-contained. Let me, therefore, give a pseudohistorical introduction to the problem with a piece of science fiction (or fictional science).

In the days before the advent of microscopes, there lived a Pasteur-like scientist whom, following the usual custom, I shall call Jones. Reflecting on the fact that certain diseases seemed to be transmitted from one person to another by means of bodily contact or by contact with articles handled previously by an afflicted person, Jones began to speculate about the mechanism of the transmission. As a "heuristic crutch," he recalled that there is an obvious *observable* mechanism for transmission of certain afflictions (such as body lice), and he postulated that all, or most, infectious diseases were spread in a similar manner but that in most cases the corresponding "bugs" were too small to be seen and, possibly, that some of them lived inside the bodies of their hosts. Jones proceeded to develop his theory and to examine its testable consequences. Some of these seemed to be of great importance for preventing the spread of disease.

After years of struggle with incredulous recalcitrance, Jones managed to get some of his preventative measures adopted. Contact with or proximity to diseased persons was avoided when possible, and articles which they handled were "disinfected" (a word coined by Jones) either by means of high temperatures or by treating them with certain toxic preparations which Jones termed "disinfectants." The results were spectacular: within ten years the death rate had declined 40 per cent. Jones and his theory received their well-deserved recognition.

However, the "crobes" (the theoretical term coined by Jones to refer to the disease-producing organisms) aroused considerable anxiety among many of the philosophers and philosophically inclined scientists of the day. The expression of this anxiety usually began something like this: "In order to account for the facts, Jones must assume that his crobes are too small to be seen. Thus the very postulates of his theory preclude their being observed; they are *unobservable in principle*." (Recall that no one had envisaged such a thing as a microscope.) This common prefatory remark was then followed by a number of different "analyses" and "interpretations" of Jones' theory. According to one of these, the tiny organisms were merely convenient fictions—*façons de parler*—extremely useful as heuristic devices for facilitating (in the "context of discovery") the thinking

of scientists but not to be taken seriously in the sphere of cognitive knowledge (in the "context of justification"). A closely related view was that Jones' theory was merely an instrument, useful for organizing observation statements and (thus) for producing desired results, and that, therefore, it made no more sense to ask what was the nature of the entities to which it referred than it did to ask what was the nature of the entities to which a hammer or any other tool referred.³ "Yes," a philosopher might have said, "Jones' theoretical expressions are just meaningless sounds or marks on paper which, when correlated with observation sentences by appropriate syntactical rules, enable us to predict successfully and otherwise organize data in a convenient fashion." These philosophers called themselves "instrumentalists."

According to another view (which, however, soon became unfashionable), although expressions containing Jones' theoretical terms were genuine sentences, they were translatable without remainder into a set (perhaps infinite) of observation sentences. For example, 'There are crobes of disease X on this article' was said to translate into something like this: 'If a person handles this article without taking certain precautions, he will (probably) contract disease X; and if this article is first raised to a high temperature, then if a person handles it at any time afterward, before it comes into contact with another person with disease X, he will (probably) not contract disease X; and . . .'

Now virtually all who held any of the views so far noted granted, even insisted, that theories played a useful and legitimate role in the scientific enterprise. Their concern was the elimination of "pseudo problems" which might arise, say, when one began wondering about the "reality of supraempirical entities," etc. However, there was also a school of thought, founded by a psychologist named Pelter, which differed in an interesting manner from such positions as these. Its members held that while Jones' crobes might very well exist and enjoy "full-blown reality," they should not be the concern of medical research at all. They insisted that if Jones had employed the correct methodology, he would have discovered, even sooner and with much less effort, all of the observation laws relating to disease contraction, transmission, etc. without introducing superfluous links (the crobes) into the causal chain.

Now, lest any reader find himself waxing impatient, let me hasten to emphasize that this crude parody is not intended to convince anyone, or even to cast serious doubt upon sophisticated varieties of any of the reductionistic positions caricatured (some of them not too severely, I would contend) above. I am well aware that there are theoretical entities and theoretical entities, some of whose conceptual and theoretical statuses differ in important respects from Jones' crobes. (I shall discuss some of these later.) Allow me, then, to bring the Jonesean prelude to our examination of observability to a hasty conclusion.

Now Jones had the good fortune to live to see the invention of the

compound microscope. His crobes were “observed” in great detail, and it became possible to identify the specific kind of *microbe* (for so they began to be called) which was responsible for each different disease. Some philosophers freely admitted error and were converted to realist positions concerning theories. Others resorted to subjective idealism or to a thoroughgoing phenomenism, of which there were two principal varieties. According to one, the one “legitimate” observation language had for its descriptive terms only those which referred to sense data. The other maintained the stronger thesis that *all* “factual” statements were *translatable* without remainder into the sense-datum language. In either case, any two non-sense data (e.g., a theoretical entity and what would ordinarily be called an “observable physical object”) had virtually the same status. Others contrived means of modifying their views much less drastically. One group maintained that Jones’ crobes actually never had been unobservable in principle, for, they said, the theory did not imply the impossibility of finding a means (e.g., the microscope) of observing them. A more radical contention was that the crobes were not observed at all; it was argued that what was seen by means of the microscope was just a shadow or an image rather than a corporeal organism.

■ | The Observational-Theoretical Dichotomy

Let us turn from these fictional philosophical positions and consider some of the actual ones to which they roughly correspond. Taking the last one first, it is interesting to note the following passage from Bergmann: “But it is only fair to point out that if this . . . methodological and terminological analysis [for the thesis that there are no atoms] . . . is strictly adhered to, even stars and microscopic objects are not physical things in a literal sense, but merely by courtesy of language and pictorial imagination. This might seem awkward. But when I look through a microscope, all I see is a patch of color which creeps through the field like a shadow over a wall. And a shadow, though real, is certainly not a physical thing.”⁴

I should like to point out that it is also the case that if this analysis is strictly adhered to, we cannot observe physical things through opera glasses, or even through ordinary spectacles, and one begins to wonder about the status of what we see through an ordinary windowpane. And what about distortions due to temperature gradients—however small and, thus, always present—in the ambient air? It really does “seem awkward” to say that when people who wear glasses describe what they see they are talking about shadows, while those who employ unaided vision talk about physical things—or that when we look through a windowpane, we can only *infer* that it is raining, while if we raise the window, we may “observe directly” that it is. The point I am making is that there is, in principle, a

continuous series beginning with looking through a vacuum and containing these as members: looking through a windowpane, looking through glasses, looking through binoculars, looking through a low-power microscope, looking through a high-power microscope, etc., in the order given. The important consequence is that, so far, we are left without criteria which would enable us to draw a nonarbitrary line between "observation" and "theory." Certainly, we will often find it convenient to draw such a to-some-extent-arbitrary line; but its position will vary widely from context to context. (For example, if we are determining the resolving characteristics of a certain microscope, we would certainly draw the line beyond ordinary spectacles, probably beyond simple magnifying glasses, and possibly beyond another microscope with a lower power of resolution.) But what ontological ice does a mere methodologically convenient observational-theoretical dichotomy cut? Does an entity attain physical thinghood and/or "real existence" in one context only to lose it in another? Or, we may ask, recalling the continuity from observable to unobservable, is what is seen through spectacles a "little bit less real" or does it "exist to a slightly less extent" than what is observed by unaided vision?⁵

However, it might be argued that things seen through spectacles and binoculars look like ordinary physical objects, while those seen through microscopes and telescopes look like shadows and patches of light. I can only reply that this does not seem to me to be the case, particularly when looking at the moon, or even Saturn, through a telescope or when looking at a small, though "directly observable," physical object through a low-power microscope. Thus, again, a continuity appears.

"But," it might be objected, "theory tells us that what we see by means of a microscope is a real image, which is certainly distinct from the object on the stage." Now first of all, it should be remarked that it seems odd that one who is espousing an austere empiricism which requires a sharp observational-language/theoretical-language distinction (and one in which the former language has a privileged status) should need a theory in order to tell him what is observable. But, letting this pass, what is to prevent us from saying that we still observe the object on the stage, even though a "real image" may be involved? Otherwise, we shall be strongly tempted by phenomenalist demons, and at this point we are considering a physical-object observation language rather than a sense-datum one. (Compare the traditional puzzles: Do I see one physical object or two when I punch my eyeball? Does one object split into two? Or do I see one object and one image? Etc.)

Another argument for the continuous transition from the observable to the unobservable (theoretical) may be adduced from theoretical considerations themselves. For example, contemporary valency theory tells us that there is a virtually continuous transition from very small molecules (such as those of hydrogen) through "medium-sized" ones (such as those of the fatty acids, polypeptides, proteins, and viruses) to extremely large

ones (such as crystals of the salts, diamonds, and lumps of polymeric plastic). The molecules in the last-mentioned group are macro, "directly observable" physical objects but are, nevertheless, genuine, single molecules; on the other hand, those in the first-mentioned group have the same perplexing properties as subatomic particles (de Broglie waves, Heisenberg indeterminacy, etc.). Are we to say that a large protein molecule (e.g., a virus) which can be "seen" only with an electron microscope is a little less real or exists to somewhat less an extent than does a molecule of a polymer which can be seen with an optical microscope? And does a hydrogen molecule partake of only an infinitesimal portion of existence or reality? Although there certainly *is* a continuous transition from observability to unobservability, any talk of such a continuity from full-blown existence to nonexistence is, clearly, nonsense.

Let us now consider the next to last modified position which was adopted by our fictional philosophers. According to them, it is only those entities which are *in principle* impossible to observe that present special problems. What kind of impossibility is meant here? Without going into a detailed discussion of the various types of impossibility, about which there is abundant literature with which the reader is no doubt familiar, I shall assume what usually seems to be granted by most philosophers who talk of entities which are unobservable in principle—i.e., that the theory(s) itself (coupled with a physiological theory of perception, I would add) entails that such entities are unobservable.

We should immediately note that if this analysis of the notion of unobservability (and, hence, of observability) is accepted, then its use as a means of delimiting the observation language seems to be precluded for those philosophers who regard theoretical expressions as elements of a calculating device—as meaningless strings of symbols. For suppose they wished to determine whether or not 'electron' was a theoretical term. First, they must see whether the theory entails the sentence 'Electrons are unobservable.' So far, so good, for their calculating devices are said to be able to select genuine sentences, provided they contain no theoretical terms. But what about the selected "sentence" itself? Suppose that 'electron' is an observation term. It follows that the expression is a genuine sentence and asserts that electrons are unobservable. But this entails that 'electron' is *not* an observation term. Thus if 'electron' is an observation term, then it is *not* an observation term. Therefore it is not an observation term. But then it follows that 'Electrons are unobservable' is not a genuine sentence and does not assert that electrons are unobservable, since it is a meaningless string of marks and does not assert anything whatever. Of course, it could be stipulated that when a theory "selects" a meaningless expression of the form 'Xs are unobservable,' then 'X' is to be taken as a theoretical term. But this seems rather arbitrary.

But, assuming that well-formed theoretical expressions are genuine sentences, what shall we say about unobservability in principle? I shall

begin by putting my head on the block and argue that the present-day status of, say, electrons is in many ways similar to that of Jones' crobes before microscopes were invented. I am well aware of the numerous theoretical arguments for the impossibility of observing electrons. But suppose new entities are discovered which interact with electrons in such a mild manner that if an electron is, say, in an eigenstate of position, then, in certain circumstances, the interaction does not disturb it. Suppose also that a drug is discovered which vastly alters the human perceptual apparatus—perhaps even activates latent capacities so that a new sense modality emerges. Finally, suppose that in our altered state we are able to perceive (not necessarily visually) by means of these new entities in a manner roughly analogous to that by which we now see by means of photons. To make this a little more plausible, suppose that the energy eigenstates of the electrons in some of the compounds present in the relevant perceptual organ are such that even the weak interaction with the new entities alters them and also that the cross sections, relative to the new entities, of the electrons and other particles of the gases of the air are so small that the chance of any interaction here is negligible. Then we might be able to “observe directly” the position and possibly the approximate diameter and other properties of some electrons. It would follow, of course, that quantum theory would have to be altered in some respects, since the new entities do not conform to all its principles. But however improbable this may be, it does not, I maintain, involve any logical or conceptual absurdity. Furthermore, the modification necessary for the inclusion of the new entities would not necessarily change the meaning of the term ‘electron.’⁶

Consider a somewhat less fantastic example, and one which does not involve any change in physical theory. Suppose a human mutant is born who is able to “observe” ultraviolet radiation, or even X rays, in the same way we “observe” visible light.

Now I think that it is extremely improbable that we will ever observe electrons directly (i.e., that it will ever be reasonable to assert that we have so observed them). But this is neither here nor there; it is not the purpose of this essay to predict the future development of scientific theories, and, hence, it is not its business to decide what actually is observable or what will become observable (in the more or less intuitive sense of ‘observable’ with which we are now working). After all, we are operating, here, under the assumption that it is theory, and thus science itself, which tells us what is or is not, in this sense, observable (the ‘in principle’ seems to have become superfluous). And this is the heart of the matter; for it follows that, at least for this sense of ‘observable,’ there are no a priori or philosophical criteria for separating the observable from the unobservable. By trying to show that we can talk about the *possibility* of observing electrons without committing logical or conceptual blunders, I have been trying to

support the thesis that any (nonlogical) term is a *possible* candidate for an observation term.

There is another line which may be taken in regard to delimitation of the observation language. According to it, the proper term with which to work is not 'observable' but, rather 'observed.' There immediately comes to mind the tradition beginning with Locke and Hume (No idea without a preceding impression!), running through Logical Atomism and the Principle of Acquaintance, and ending (perhaps) in contemporary positivism.* Since the numerous facets of this tradition have been extensively examined and criticized in the literature, I shall limit myself here to a few summary remarks.

Again, let us consider at this point only observation languages which contain ordinary physical-object terms (along with observation predicates, etc., of course). Now, according to this view, all descriptive terms of the observation language must refer to that which has been observed. How is this to be interpreted? Not too narrowly, presumably, otherwise each language user would have a different observation language. The name of my Aunt Mamie, of California, whom I have never seen, would not be in my observation language, nor would 'snow' be an observation term for many Floridians. One could, of course, set off the observation language by means of this awkward restriction, but then, obviously, not being the referent of an observation term would have no bearing on the ontological status of Aunt Mamie or that of snow.

Perhaps it is intended that the referents of observation terms must be members of a *kind* some of whose members have been observed or instances of a *property* some of whose instances have been observed. But there are familiar difficulties here. For example, given any entity, we can always find a kind whose only member is the entity in question; and surely expressions such as 'men over 14 feet tall' should be counted as observational even though no instances of the "property" of being a man over 14 feet tall have been observed. It would seem that this approach must soon fall back upon some notion of simples or determinables vs. determinates. But is it thereby saved? If it is held that only those terms which

* Logical atomism is the doctrine held by Bertrand Russell and Ludwig Wittgenstein in the early decades of the twentieth century, although both later abandoned it. Logical atomism holds that all meaningful propositions of any degree of complexity are truth-functional compounds of atomic propositions. Atomic propositions have no truth-functional structure and refer directly to nonlinguistic atomic facts. Russell's principle of acquaintance asserts that every proposition we can understand must be composed entirely of elements with which we are acquainted, where acquaintance is the immediate and infallible apprehension of an object. By identifying atomic facts with sense data, the logical positivists argued that acquaintance with them is necessary both for understanding what propositions mean and for having any empirical knowledge.

refer to observed simples or observed determinates are observation terms, we need only remind ourselves of such instances as Hume's notorious missing shade of blue.* And if it is contended that in order to be an observation term an expression must at least refer to an observed determinable, then we can always find such a determinable which is broad enough in scope to embrace any entity whatever. But even if these difficulties can be circumvented, we see (as we knew all along) that this approach leads inevitably into phenomenalism, which is a view with which we have not been concerning ourselves.

Now it is not the purpose of this essay to give a detailed critique of phenomenalism. For the most part, I simply assume that it is untenable, at least in any of its translatability varieties.⁷ However, if there are any unreconstructed phenomenalists among the readers, my purpose, insofar as they are concerned, will have been largely achieved if they will grant what I suppose most of them would stoutly maintain anyway, i.e., that theoretical entities are no worse off than so-called observable physical objects.

Nevertheless, a few considerations concerning phenomenalism and related matters may cast some light upon the observational-theoretical dichotomy and, perhaps, upon the nature of the "observation language." As a preface, allow me some overdue remarks on the latter. Although I have contended that the line between the observable and the unobservable is diffuse, that it shifts from one scientific problem to another, and that it is constantly being pushed toward the "unobservable" end of the spectrum as we develop better means of observation—better instruments—it would, nevertheless, be fatuous to minimize the importance of the observation base, for it is absolutely necessary as a confirmation base for statements which do refer to entities which are unobservable at a given time. But we should take as its basis and its unit not the "observational term" but, rather, the quickly decidable sentence. (I am indebted to Feyerabend, *loc. cit.*, for this terminology.) A quickly decidable sentence (in the technical sense employed here) may be defined as a singular, nonanalytic sentence such that a reliable, reasonably sophisticated language user can very quickly decide⁸ whether to assert it or deny it when he is reporting on an occurrent situation. 'Observation term' may now be defined as a 'descriptive (non-logical) term which may occur in a quickly decidable sentence,' and 'observation sentence' as a 'sentence whose only descriptive terms are observation terms.'

Returning to phenomenalism, let me emphasize that I am not among

* Hume allowed that we can form the idea of a particular shade of blue even though we have never seen it by filling in the gap (so to speak) between the shades of blue that we have perceived. This apparently contradicts Hume's empiricist principle that every simple idea is copied from a preceding simple impression. See David Hume, *An Enquiry Concerning Human Understanding*, Section II.

those philosophers who hold that there are no such things as sense contents (even sense data), nor do I believe that they play no important role in our perception of "reality." But the fact remains that the referents of most (not all) of the statements of the linguistic framework used in everyday life and in science are *not* sense contents but, rather, physical objects and other publicly observable entities. Except for pains, odors, "inner states," etc., *we do not usually observe sense contents*; and although there is good reason to believe that they play an indispensable role in observation, *we are usually not aware of them when we* (visually or tactilely) *observe physical objects*. For example, when I observe a distorted, obliquely reflected image in a mirror, I may seem to be seeing a baby elephant standing on its head; later I discover it is an image of Uncle Charles taking a nap with his mouth open and his hand in a peculiar position. Or, passing my neighbor's home at a high rate of speed, I observe that he is washing a car. If asked to report these observations I could quickly and easily report a baby elephant and a washing of a car; I probably would not, without subsequent observations, be able to report what colors, shapes, etc. (i.e., what sense data) were involved.

Two questions naturally arise at this point. How is it that we can (sometimes) quickly decide the truth or falsity of a pertinent observation sentence? and, What role do sense contents play in the appropriate tokening of such sentences? The heart of the matter is that these are primarily scientific-theoretical questions rather than "purely logical," "purely conceptual," or "purely epistemological." If theoretical physics, psychology, neurophysiology, etc., were sufficiently advanced, we could give satisfactory answers to these questions, using, in all likelihood, the physical-thing language as our observation language and *treating sensations, sense contents, sense data, and "inner states" as theoretical* (yes, theoretical!) *entities*.⁹

It is interesting and important to note that, even before we give completely satisfactory answers to the two questions considered above, we can, with due effort and reflection, train ourselves to "observe directly" what were once theoretical entities—the sense contents (color sensations, etc.)—involved in our perception of physical things. As has been pointed out before, we can also come to observe other kinds of entities which were once theoretical. Those which most readily come to mind involve the use of instruments as aids to observation. Indeed, using our painfully acquired theoretical knowledge of the world, we come to see that we "directly observe" many kinds of so-called theoretical things. After listening to a dull speech while sitting on a hard bench, we begin to become poignantly aware of the presence of a considerably strong gravitational field, and as Professor Feyerabend is fond of pointing out, if we were carrying a heavy suitcase in a changing gravitational field, we could observe the changes of the $G_{\mu\nu}$ of the metric tensor.

I conclude that our drawing of the observational-theoretical line at

any given point is an accident and a function of our physiological makeup, our current state of knowledge, and the instruments we happen to have available and, therefore, that it has no ontological significance whatever. . . .

■ | Notes

1. E. Nagel, *The Structure of Science* (New York: Harcourt, Brace, and World, 1961), Ch. 6.
2. For the genesis and part of the content of some of the ideas expressed herein, I am indebted to a number of sources; some of the more influential are H. Feigl, "Existential Hypotheses," *Philosophy of Science*, 17:35–62 (1950); P. K. Feyerabend, "An Attempt at a Realistic Interpretation of Experience," *Proceedings of the Aristotelian Society*, 58:144–170 (1958); N. R. Hanson, *Patterns of Discovery* (Cambridge: Cambridge University Press, 1958); E. Nagel, *loc. cit.*; Karl Popper, *The Logic of Scientific Discovery* (London: Hutchinson, 1959); M. Scriven, "Definitions, Explanations, and Theories," in *Minnesota Studies in the Philosophy of Science*, Vol. II, H. Feigl, M. Scriven, and G. Maxwell, eds. (Minneapolis: University of Minnesota Press, 1958); Wilfrid Sellars, "Empiricism and the Philosophy of Mind," in *Minnesota Studies in the Philosophy of Science*, Vol. I, H. Feigl and M. Scriven, eds. (Minneapolis: University of Minnesota Press, 1956), and "The Language of Theories," in *Current Issues in the Philosophy of Science*, H. Feigl and G. Maxwell, eds. (New York: Holt, Rinehart, and Winston, 1961).
3. I have borrowed the hammer analogy from E. Nagel, "Science and [Feigl's] Semantic Realism," *Philosophy of Science*, 17:174–181 (1950), but it should be pointed out that Professor Nagel makes it clear that he does not necessarily subscribe to the view which he is explaining.
4. G. Bergmann, "Outline of an Empiricist Philosophy of Physics," *American Journal of Physics*, 11:248–258; 335–342 (1943), reprinted in *Readings in the Philosophy of Science*, H. Feigl and M. Brodbeck, eds. (New York: Appleton-Century-Crofts, 1953), pp. 262–287.
5. I am not attributing to Professor Bergmann the absurd views suggested by these questions. He seems to take a sense-datum language as his observation language (the base of what he called "the empirical hierarchy"), and, in some ways, such a position is more difficult to refute than one which purports to take an "observable physical-object" view. However, I believe that demolishing the straw men with which I am now dealing amounts to desirable preliminary "therapy." Some non-realist interpretations of theories which embody the presupposition that the observable-theoretical distinction is sharp and ontologically crucial seem to me to entail positions which correspond to such straw men rather closely.
6. For arguments that it is possible to alter a theory without altering the meanings of its terms, see my "Meaning Postulates in Scientific Theories," in *Current Issues in the Philosophy of Science*, Feigl and Maxwell, eds.

7. The reader is no doubt familiar with the abundant literature concerned with this issue. See, for example, Sellars' "Empiricism and the Philosophy of Mind," which also contains references to other pertinent works.
8. We may say "noninferentially" decide, provided this is interpreted liberally enough to avoid starting the entire controversy about observability all over again.
9. Cf. Sellars, "Empiricism and the Philosophy of Mind." As Professor Sellars points out, this is the crux of the "other-minds" problem. Sensations and inner states (relative to an intersubjective observation language, I would add) are theoretical entities (and they "really exist") and *not* merely actual and/or possible behavior. Surely it is the unwillingness to countenance theoretical entities—the hope that every sentence is translatable not only into some observation language but into the physical-thing language—which is responsible for the "logical behaviorism" of the neo-Wittgensteinians.

Arguments Concerning Scientific Realism

The rigour of science requires that we distinguish well the undraped figure of nature itself from the gay-coloured vesture with which we clothe it at our pleasure.

—Heinrich Hertz, quoted by Ludwig Boltzmann,
letter to *Nature*, 28 February 1895

In our century, the first dominant philosophy of science was developed as part of logical positivism. Even today, such an expression as ‘the received view of theories’ refers to the views developed by the logical positivists, although their heyday preceded the Second World War.

In this chapter I shall examine, and criticize, the main arguments that have been offered for scientific realism. These arguments occurred frequently as part of a critique of logical positivism. But it is surely fair to discuss them in isolation, for even if scientific realism is most easily understood as a reaction against positivism, it should be able to stand alone. The alternative view which I advocate—for lack of a traditional name I shall call it *constructive empiricism*—is equally at odds with positivist doctrine.

1 | Scientific Realism and Constructive Empiricism

In philosophy of science, the term ‘scientific realism’ denotes a precise position on the question of how a scientific theory is to be understood, and what scientific activity really is. I shall attempt to define this position, and to canvass its possible alternatives. Then I shall indicate, roughly and briefly, the specific alternative which I shall advocate. . . .

FROM Bas C. van Fraassen, *The Scientific Image* (Oxford: Clarendon Press, 1980), 6–21, 23–25, 31–40.

1.1 STATEMENT OF SCIENTIFIC REALISM

What exactly is scientific realism? A naïve statement of the position would be this: the picture which science gives us of the world is a true one, faithful in its details, and the entities postulated in science really exist: the advances of science are discoveries, not inventions. That statement is too naïve; it attributes to the scientific realist the belief that today's theories are correct. It would mean that the philosophical position of an earlier scientific realist such as C. S. Peirce had been refuted by empirical findings. I do not suppose that scientific realists wish to be committed, as such, even to the claim that science will arrive in due time at theories true in all respects—for the growth of science might be an endless self-correction; or worse, Armageddon might occur too soon.

But the naïve statement has the right flavour. It answers two main questions: it characterizes a scientific theory as a story about what there really is, and scientific activity as an enterprise of discovery, as opposed to invention. The two questions of what a scientific theory is, and what a scientific theory does, must be answered by any philosophy of science. The task we have at this point is to find a statement of scientific realism that shares these features with the naïve statement, but does not saddle the realists with unacceptably strong consequences. It is especially important to make the statement as weak as possible if we wish to argue against it, so as not to charge at windmills.

As clues I shall cite some passages most of which will also be examined below in the contexts of the authors' arguments. A statement of Wilfrid Sellars is this:

to have good reason for holding a theory is *ipso facto* to have good reason for holding that the entities postulated by the theory exist.¹

This addresses a question of epistemology, but also throws some indirect light on what it is, in Sellars's opinion, to hold a theory. Brian Ellis, who calls himself a scientific entity realist rather than a scientific realist, appears to agree with that statement of Sellars, but gives the following formulation of a stronger view:

I understand scientific realism to be the view that the theoretical statements of science are, or purport to be, true generalized descriptions of reality.²

This formulation has two advantages: It focuses on the understanding of the theories without reference to reasons for belief, and it avoids the suggestion that to be a realist you must believe current scientific theories to be true. But it gains the latter advantage by use of the word 'purport', which may generate its own puzzles.

Hilary Putnam, in a passage which I shall cite again in Section 7, gives a formulation which he says he learned from Michael Dummett:

A realist (with respect to a given theory or discourse) holds that (1) the sentences of that theory are true or false; and (2) that what makes them true or false is something external—that is to say, it is not (in general) our sense data, actual or potential, or the structure of our minds, or our language, etc.¹

He follows this soon afterwards with a further formulation which he credits to Richard Boyd:

That terms in mature scientific theories typically refer (this formulation is due to Richard Boyd), that the theories accepted in a mature science are typically approximately true, that the same term can refer to the same thing even when it occurs in different theories—these statements are viewed by the scientific realist . . . as part of any adequate scientific description of science and its relations to its objects.²

None of these were intended as definitions. But they show I think that truth must play an important role in the formulation of the basic realist position. They also show that the formulation must incorporate an answer to the question what it is to *accept* or *hold* a theory. I shall now propose such a formulation, which seems to me to make sense of the above remarks, and also renders intelligible the reasoning by realists which I shall examine below—without burdening them with more than the minimum required for this.

Science aims to give us, in its theories, a literally true story of what the world is like; and acceptance of a scientific theory involves the belief that it is true. This is the correct statement of scientific realism.

Let me defend this formulation by showing that it is quite minimal, and can be agreed to by anyone who considers himself a scientific realist. The naïve statement said that science tells a true story; the correct statement says only that it is the aim of science to do so. The aim of science is of course not to be identified with individual scientists' motives. The aim of the game of chess is to checkmate your opponent; but the motive for playing may be fame, gold, and glory. What the aim is determines what counts as success in the enterprise as such; and this aim may be pursued for any number of reasons. Also, in calling something *the* aim, I do not deny that there are other subsidiary aims which may or may not be means to that end: everyone will readily agree that simplicity, informativeness, predictive power, explanation are (also) virtues. Perhaps my formulation can even be accepted by any philosopher who considers the most important aim of science to be something which only *requires* the finding of true theories—given that I wish to give the weakest formulation of the doctrine that is generally acceptable.

I have added 'literally' to rule out as realist such positions as imply that science is true if 'properly understood' but literally false or meaningless. For that would be consistent with conventionalism, logical positivism, and instrumentalism. I will say more about this below; and also in Section 7 where I shall consider Dummett's views further.

The second part of the statement touches on epistemology. But it only equates acceptance of a theory with belief in its truth.⁵ It does not imply that anyone is ever rationally warranted in forming such a belief. We have to make room for the epistemological position, today the subject of considerable debate, that a rational person never assigns personal probability 1 to any proposition except a tautology. It would, I think, be rare for a scientific realist to take this stand in epistemology, but it is certainly possible.⁶

To understand qualified acceptance we must first understand acceptance *tout court*. If acceptance of a theory involves the belief that it is true, then tentative acceptance involves the tentative adoption of the belief that it is true. If belief comes in degrees, so does acceptance, and we may then speak of a degree of acceptance involving a certain degree of belief that the theory is true. This must of course be distinguished from belief that the theory is approximately true, which seems to mean belief that some member of a class centring on the mentioned theory is (exactly) true. In this way the proposed formulation of realism can be used regardless of one's epistemological persuasion.

1.2 ALTERNATIVES TO REALISM

Scientific realism is the position that scientific theory construction aims to give us a literally true story of what the world is like, and that acceptance of a scientific theory involves the belief that it is true. Accordingly, anti-realism is a position according to which the aim of science can well be served without giving such a literally true story, and acceptance of a theory may properly involve something less (or other) than belief that it is true.

What does a scientist do then, according to these different positions? According to the realist, when someone proposes a theory, he is asserting it to be true. But according to the anti-realist, the proposer does not assert the theory; *he displays it*, and claims certain virtues for it. These virtues may fall short of truth: empirical adequacy, perhaps; comprehensiveness, acceptability for various purposes. This will have to be spelt out, for the details here are not determined by the denial of realism. For now we must concentrate on the key notions that allow the generic division.

The idea of a literally true account has two aspects: the language is to be literally construed; and so construed, the account is true. This divides the anti-realists into two sorts. The first sort holds that science is or aims to be true, properly (but not literally) construed. The second holds that the language of science should be literally construed, but its theories need

not be true to be good. The anti-realism I shall advocate belongs to the second sort.

It is not so easy to say what is meant by a literal construal. The idea comes perhaps from theology, where fundamentalists construe the Bible literally, and liberals have a variety of allegorical, metaphorical, and analogical interpretations, which 'demythologize'. The problem of explicating 'literal construal' belongs to the philosophy of language. In Section 7 below, where I briefly examine some of Michael Dummett's views, I shall emphasize that 'literal' does not mean 'truth-valued'. The term 'literal' is well enough understood for general philosophical use, but if we try to explicate it we find ourselves in the midst of the problem of giving an adequate account of natural language. It would be bad tactics to link an inquiry into science to a commitment to some solution to that problem. The following remarks, and those in Section 7, should fix the usage of 'literal' sufficiently for present purposes.

The decision to rule out all but literal construals of the language of science, rules out those forms of anti-realism known as *positivism* and *instrumentalism*. First, on a literal construal, the apparent statements of science really are statements, *capable of* being true or false. Secondly, although a literal construal can elaborate, it cannot change logical relationships. (It is possible to elaborate, for instance, by identifying what the terms designate. The 'reduction' of the language of phenomenological thermodynamics to that of statistical mechanics is like that: bodies of gas are identified as aggregates of molecules, temperature as mean kinetic energy, and so on.) On the positivists' interpretation of science, theoretical terms have meaning only through their connection with the observable. Hence they hold that two theories may in fact *say the same thing* although in form they contradict each other. (Perhaps the one says that all matter consists of atoms, while the other postulates instead a universal continuous medium; they will say the same thing nevertheless if they agree in their observable consequences, according to the positivists.) But two theories which contradict each other in such a way can 'really' be saying the same thing only if they are not literally construed. Most specifically, if a theory says that something exists, then a literal construal may elaborate on what that something is, but will not remove the implication of existence.

There have been many critiques of positivist interpretations of science, and there is no need to repeat them. . . .

1.3 CONSTRUCTIVE EMPIRICISM

To insist on a literal construal of the language of science is to rule out the construal of a theory as a metaphor or simile, or as intelligible only after it is 'demythologized' or subjected to some other sort of 'translation' that does not preserve logical form. If the theory's statements include 'There are electrons', then the theory says that there are electrons. If in

addition they include 'Electrons are not planets', then the theory says, in part, that there are entities other than planets.

But this does not settle very much. It is often not at all obvious whether a theoretical term refers to a concrete entity or a mathematical entity. Perhaps one tenable interpretation of classical physics is that there are no concrete entities which are forces—that 'there are forces such that . . .' can always be understood as a mathematical statement asserting the existence of certain functions. That is debatable.

Not every philosophical position concerning science which insists on a literal construal of the language of science is a realist position. For this insistence relates not at all to our epistemic attitudes toward theories, nor to the aim we pursue in constructing theories, but only to the correct understanding of *what a theory says*. (The fundamentalist theist, the agnostic, and the atheist presumably agree with each other (though not with liberal theologians) in their understanding of the statement that God, or gods, or angels exist.) After deciding that the language of science must be literally understood, we can still say that there is no need to believe good theories to be true, nor to believe *ipso facto* that the entities they postulate are real.

Science aims to give us theories which are empirically adequate; and acceptance of a theory involves as belief only that it is empirically adequate. This is the statement of the anti-realist position I advocate; I shall call it *constructive empiricism*.

This formulation is subject to the same qualifying remarks as that of scientific realism in Section 1.1 above. In addition it requires an explication of 'empirically adequate'. For now, I shall leave that with the preliminary explication that a theory is empirically adequate exactly if what it says about the observable things and events in this world, is true—exactly if it 'saves the phenomena'. A little more precisely: such a theory has at least one model that all the actual phenomena fit inside. I must emphasize that this refers to *all* the phenomena; these are not exhausted by those actually observed, nor even by those observed at some time, whether past, present, or future. . . .

The distinction I have drawn between realism and anti-realism, in so far as it pertains to acceptance, concerns only how much belief is involved therein. Acceptance of theories (whether full, tentative, to a degree, etc.) is a phenomenon of scientific activity which clearly involves more than belief. One main reason for this is that we are never confronted with a complete theory. So if a scientist accepts a theory, he thereby involves himself in a certain sort of research programme. That programme could well be different from the one acceptance of another theory would have given him, even if those two (very incomplete) theories are equivalent to each other with respect to everything that is observable—in so far as they go.

Thus acceptance involves not only belief but a certain commitment.

Even for those of us who are not working scientists, the acceptance involves a commitment to confront any future phenomena by means of the conceptual resources of this theory. It determines the terms in which we shall seek explanations. If the acceptance is at all strong, it is exhibited in the person's assumption of the role of explainer, in his willingness to answer questions *ex cathedra*. Even if you do not accept a theory, you can engage in discourse in a context in which language use is guided by that theory—but acceptance produces such contexts. There are similarities in all of this to ideological commitment. A commitment is of course not true or false: The confidence exhibited is that it will be *vindicated*.

This is a preliminary sketch of the *pragmatic* dimension of theory acceptance. Unlike the epistemic dimension, it does not figure overtly in the disagreement between realist and anti-realist. But because the amount of belief involved in acceptance is typically less according to anti-realists, they will tend to make more of the pragmatic aspects. It is as well to note here the important difference. Belief that a theory is true, or that it is empirically adequate, does not imply, and is not implied by, belief that full acceptance of the theory will be vindicated. To see this, you need only consider here a person who has quite definite beliefs about the future of the human race, or about the scientific community and the influences thereon and practical limitations we have. It might well be, for instance, that a theory which is empirically adequate will not combine easily with some other theories which we have accepted in fact, or that Armageddon will occur before we succeed. Whether belief that a theory is true, or that it is empirically adequate, can be equated with belief that acceptance of it would, under ideal research conditions, be vindicated in the long run, is another question. It seems to me an irrelevant question within philosophy of science, because an affirmative answer would not obliterate the distinction we have already established by the preceding remarks. (The question may also assume that counterfactual statements are objectively true or false, which I would deny.)

Although it seems to me that realists and anti-realists need not disagree about the pragmatic aspects of theory acceptance, I have mentioned it here because I think that typically they do. We shall find ourselves returning time and again, for example, to requests for explanation to which realists typically attach an objective validity which anti-realists cannot grant.

2 | The Theory/Observation 'Dichotomy'

For good reasons, logical positivism dominated the philosophy of science for thirty years. In 1960, the first volume of *Minnesota Studies in the Philosophy of Science* published Rudolf Carnap's 'The Methodological

Status of Theoretical Concepts', which is, in many ways, the culmination of the positivist programme. It interprets science by relating it to an observation language (a postulated part of natural language which is devoid of theoretical terms). Two years later this article was followed in the same series by Grover Maxwell's 'The Ontological Status of Theoretical Entities', in title and theme a direct counter to Carnap's. This is the *locus classicus* for the new realists' contention that the theory/observation distinction cannot be drawn.

I shall examine some of Maxwell's points directly, but first a general remark about the issue. Such expressions as 'theoretical entity' and 'observable-theoretical dichotomy' are, on the face of it, examples of category mistakes. Terms or concepts are theoretical (introduced or adapted for the purposes of theory construction); entities are observable or unobservable. This may seem a little point, but it separates the discussion into two issues. Can we divide our language into a theoretical and non-theoretical part? On the other hand, can we classify objects and events into observable and unobservable ones?

Maxwell answers both questions in the negative, while not distinguishing them too carefully. On the first, where he can draw on well-known supportive essays by Wilfrid Sellars and Paul Feyerabend, I am in total agreement. All our language is thoroughly theory-infected. If we could cleanse our language of theory-laden terms, beginning with the recently introduced ones like 'VHF receiver', continuing through 'mass' and 'impulse' to 'element' and so on into the prehistory of language formation, we would end up with nothing useful. The way we talk, and scientists talk, is guided by the pictures provided by previously accepted theories. This is true also, as Duhem already emphasized, of experimental reports. Hygienic reconstructions of language such as the positivists envisaged are simply not on. . . .

But does this mean that we must be scientific realists? We surely have more tolerance of ambiguity than that. The fact that we let our language be guided by a given picture, at some point, does not show how much we believe about that picture. When we speak of the sun coming up in the morning and setting at night, we are guided by a picture now explicitly disavowed. When Milton wrote *Paradise Lost* he deliberately let the old geocentric astronomy guide his poem, although various remarks in passing clearly reveal his interest in the new astronomical discoveries and speculations of his time. These are extreme examples, but show that no immediate conclusions can be drawn from the theory-ladenness of our language.

However, Maxwell's main arguments are directed against the observable-unobservable distinction. Let us first be clear on what this distinction was supposed to be. The term 'observable' classifies putative entities (entities which may or may not exist). A flying horse is observable—that is why we are so sure that there aren't any—and the number

seventeen is not. There is supposed to be a correlate classification of human acts: an unaided act of perception, for instance, is an observation. A calculation of the mass of a particle from the deflection of its trajectory in a known force field, is not an observation of that mass.

It is also important here not to confuse *observing* (an entity, such as a thing, event, or process) and *observing that* (something or other is the case). Suppose one of the Stone Age people recently found in the Philippines is shown a tennis ball or a car crash. From his behaviour, we see that he has noticed them; for example, he picks up the ball and throws it. But he has not seen *that* it is a tennis ball, or *that* some event is a car crash, for he does not even have those concepts. He cannot get that information through perception; he would first have to learn a great deal. To say that he does not see the same things and events as we do, however, is just silly; it is a pun which trades on the ambiguity between seeing and seeing that. (The truth-conditions for our statement 'x observes *that* A' must be such that what concepts x has, presumably related to the language x speaks if he is human, enter as a variable into the correct truth definition, in some way. To say that x observed the tennis ball, therefore, does not imply at all that x observed that it was a tennis ball; that would require some conceptual awareness of the game of tennis.)

The arguments Maxwell gives about observability are of two sorts: one directed against the possibility of drawing such distinctions, the other against the importance that could attach to distinctions that can be drawn.

The first argument is from the continuum of cases that lie between direct observation and inference:

There is, in principle, a continuous series beginning with looking through a vacuum and containing these as members: looking through a windowpane, looking through glasses, looking through binoculars, looking through a low-power microscope, looking through a high-power microscope, etc., in the order given. The important consequence is that, so far, we are left without criteria which would enable us to draw a non-arbitrary line between 'observation' and 'theory'.⁷

This continuous series of supposed acts of observation does not correspond directly to a continuum in what is supposed observable. For if something can be seen through a window, it can also be seen with the window raised. Similarly, the moons of Jupiter can be seen through a telescope; but they can also be seen without a telescope if you are close enough. That something is observable does not automatically imply that the conditions are right for observing it now. The principle is:

X is observable if there are circumstances which are such that, if X is present to us under those circumstances, then we observe it.

This is not meant as a definition, but only as a rough guide to the avoidance of fallacies.

We may still be able to find a continuum in what is supposed detectable: perhaps some things can only be detected with the aid of an optical microscope, at least; perhaps some require an electron microscope, and so on. Maxwell's problem is: where shall we draw the line between what is observable and what is only detectable in some more roundabout way?

Granted that we cannot answer this question without arbitrariness, what follows? That 'observable' is a *vague predicate*. There are many puzzles about vague predicates, and many sophisms designed to show that, in the presence of vagueness, no distinction can be drawn at all. In Sextus Empiricus, we find the argument that incest is not immoral, for touching your mother's big toe with your little finger is not immoral, and all the rest differs only by degree. But predicates in natural language are almost all vague, and there is no problem in their use; only in formulating the logic that governs them.⁸ A vague predicate is usable provided it has clear cases and clear counter-cases. Seeing with the unaided eye is a clear case of observation. Is Maxwell then perhaps challenging us to present a clear counter-case? Perhaps so, for he says 'I have been trying to support the thesis that any (non-logical) term is a *possible* candidate for an observation term.'

A look through a telescope at the moons of Jupiter seems to me a clear case of observation, since astronauts will no doubt be able to see them as well from close up. But the purported observation of micro-particles in a cloud chamber seems to me a clearly different case—if our theory about what happens there is right. The theory says that if a charged particle traverses a chamber filled with saturated vapour, some atoms in the neighbourhood of its path are ionized. If this vapour is decompressed, and hence becomes super-saturated, it condenses in droplets on the ions, thus marking the path of the particle. The resulting silver-grey line is similar (physically as well as in appearance) to the vapour trail left in the sky when a jet passes. Suppose I point to such a trail and say: 'Look, there is a jet!'; might you not say: 'I see the vapour trail, but where is the jet?' Then I would answer: 'Look just a bit ahead of the trail . . . there! Do you see it?' Now, in the case of the cloud chamber this response is not possible. So while the particle is detected by means of the cloud chamber, and the detection is based on observation, it is clearly not a case of the particle's being observed.

As second argument, Maxwell directs our attention to the 'can' in 'what is observable is what can be observed.' An object might of course be temporarily unobservable—in a rather different sense: it cannot be observed in the circumstances in which it actually is at the moment, but could be observed if the circumstances were more favourable. In just the same way, I might be temporarily invulnerable or invisible. So we should

concentrate on 'observable' *tout court*, or on (as he prefers to say) 'unobservable in principle'. This Maxwell explains as meaning that the relevant scientific theory *entails* that the entities cannot be observed in any circumstances. But this never happens, he says, because the different circumstances could be ones in which we have different sense organs—electron-microscope eyes, for instance.

This strikes me as a trick, a change in the subject of discussion. I have a mortar and pestle made of copper and weighing about a kilo. Should I call it breakable because a giant could break it? Should I call the Empire State Building portable? Is there no distinction between a portable and a console record player? The human organism is, from the point of view of physics, a certain kind of measuring apparatus. As such it has certain inherent limitations—which will be described in detail in the final physics and biology. It is these limitations to which the 'able' in 'observable' refers—our limitations, *qua* human beings.

As I mentioned, however, Maxwell's article also contains a different sort of argument: even if there is a feasible observable/unobservable distinction, this distinction has no importance. The point at issue for the realist is, after all, the reality of the entities postulated in science. Suppose that these entities could be classified into observables and others; what relevance should that have to the question of their existence?

Logically, none. For the term 'observable' classifies putative entities, and has logically nothing to do with existence. But Maxwell must have more in mind when he says: 'I conclude that the drawing of the observational-theoretical line at any given point is an accident and a function of our physiological make-up, . . . and, therefore, that it has no ontological significance whatever.'⁹ No ontological significance if the question is only whether 'observable' and 'exists' imply each other—for they do not; but significance for the question of scientific realism?

Recall that I defined scientific realism in terms of the aim of science, and epistemic attitudes. The question is what aim scientific activity has, and how much we shall believe when we accept a scientific theory. What is the proper form of acceptance: belief that the theory, as a whole, is true; or something else? To this question, what is observable by us seems eminently relevant. Indeed, we may attempt an answer at this point: to accept a theory is (for us) to believe that it is empirically adequate—that what the theory says *about what is observable* (by us) is true.

It will be objected at once that, on this proposal, what the anti-realist decides to believe about the world will depend in part on what he believes to be his, or rather the epistemic community's, accessible range of evidence. At present, we count the human race as the epistemic community to which we belong; but this race may mutate, or that community may be increased by adding other animals (terrestrial or extra-terrestrial) through relevant ideological or moral decisions ('to count them as per-

sons'). Hence the anti-realist would, on my proposal, have to accept conditions of the form

If the epistemic community changes in fashion Y, then my beliefs about the world will change in manner Z.

To see this as an objection to anti-realism is to voice the requirement that our epistemic policies should give the same results independent of our beliefs about the range of evidence accessible to us. That requirement seems to me in no way rationally compelling; it could be honoured, I should think, only through a thoroughgoing scepticism or through a commitment to wholesale leaps of faith. But we cannot settle the major questions of epistemology *en passant* in philosophy of science; so I shall just conclude that it is, on the face of it, not irrational to commit oneself only to a search for theories that are empirically adequate, ones whose models fit the observable phenomena, while recognizing that what counts as an observable phenomenon is a function of what the epistemic community is (that *observable* is *observable-to-us*).

The notion of empirical adequacy* in this answer will have to be spelt out very carefully if it is not to bite the dust among hackneyed objections. . . . But the point stands: even if observability has nothing to do with existence (is, indeed, too anthropocentric for that), it may still have much to do with the proper epistemic attitude to science.

3 | Inference to the Best Explanation

A view advanced in different ways by Wilfrid Sellars, J. J. C. Smart, and Gilbert Harman is that the canons of rational inference require scientific realism. If we are to follow the same patterns of inference with respect to this issue as we do in science itself, we shall find ourselves irrational unless we assert the truth of the scientific theories we accept. Thus Sellars says: 'As I see it, to have good reason for holding a theory is *ipso facto* to have good reason for holding that the entities postulated by the theory exist.'¹⁰

The main rule of inference invoked in arguments of this sort is the rule of *inference to the best explanation*. The idea is perhaps to be credited to C. S. Peirce,¹¹ but the main recent attempts to explain this rule and its uses have been made by Gilbert Harman.¹² I shall only present a simplified version. Let us suppose that we have evidence *E*, and are considering several hypotheses, say *H* and *H'*. The rule then says that we should infer

* Van Fraassen offers an analysis of empirical adequacy in chapter 3 of *The Scientific Image* (1980), the book from which the present reading is excerpted.

H rather than H' exactly if H is a better explanation of E than H' is. (Various qualifications are necessary to avoid inconsistency: we should always try to move to the best over-all explanation of all available evidence.)

It is argued that we follow this rule in all 'ordinary' cases; and that if we follow it consistently everywhere, we shall be led to scientific realism, in the way Sellars's dictum suggests. And surely there are many telling 'ordinary' cases: I hear scratching in the wall, the patter of little feet at midnight, my cheese disappears—and I infer that a mouse has come to live with me. Not merely that these apparent signs of mousely presence will continue, not merely that all the observable phenomena will be as if there is a mouse; but that there really is a mouse.

Will this pattern of inference also lead us to belief in unobservable entities? Is the scientific realist simply someone who consistently follows the rules of inference that we all follow in more mundane contexts? . . .

First of all, what is meant by saying that we all *follow* a certain rule of inference? One meaning might be that we deliberately and consciously 'apply' the rule, like a student doing a logic exercise. That meaning is much too literalistic and restrictive; surely all of mankind follows the rules of logic much of the time, while only a fraction can even formulate them. A second meaning is that we act in accordance with the rules in a sense that does not require conscious deliberation. That is not so easy to make precise, since each logical rule is a rule of permission (*modus ponens* allows you to infer B from A and (if A then B), but does not forbid you to infer (B or A) instead). However, we might say that a person behaved in accordance with a set of rules in that sense if every conclusion he drew could be reached from his premisses via those rules. But this meaning is much too loose; in this sense we always behave in accordance with the rule that any conclusion may be inferred from any premiss. So it seems that to be following a rule, I must be willing to believe all conclusions it allows, while definitely unwilling to believe conclusions at variance with the ones it allows—or else, change my willingness to believe the premisses in question.

Therefore the statement that we all follow a certain rule in certain cases, is a *psychological hypothesis* about what we are willing and unwilling to do. It is an empirical hypothesis, to be confronted with data, and with rival hypotheses. Here is a rival hypothesis: we are always willing to believe that the theory which best explains the evidence, is empirically adequate (that all the observable phenomena are as the theory says they are).

In this way I can certainly account for the many instances in which a scientist appears to argue for the acceptance of a theory or hypothesis, on the basis of its explanatory success. (A number of such instances are related by Thagard.¹³) For, remember: I equate the acceptance of a scientific theory with the belief that it is empirically adequate. We have therefore two rival hypotheses concerning these instances of scientific in-

ference, and the one is apt in a realist account, the other in an anti-realist account.

Cases like the mouse in the wainscoting cannot provide telling evidence between those rival hypotheses. For the mouse *is* an observable thing; therefore 'there is a mouse in the wainscoting' and 'All observable phenomena are as if there is a mouse in the wainscoting' are totally equivalent; each implies the other (given what we know about mice).

It will be countered that it is less interesting to know whether people do follow a rule of inference than whether they ought to follow it. Granted; but the premiss that we all follow the rule of inference to the best explanation when it comes to mice and other mundane matters—that premiss is shown wanting. It is not warranted by the evidence, because that evidence is not telling *for* the premiss *as against* the alternative hypothesis I proposed, which is a relevant one in this context. . . .

4 | Limits of the Demand for Explanation

In this section and the next . . . I shall examine arguments for realism that point to explanatory power as a criterion for theory choice. That this is indeed a criterion I do not deny. But these arguments for realism succeed only if the demand for explanation is supreme—if the task of science is unfinished, *ipso facto*, as long as any pervasive regularity is left unexplained. I shall object to this line of argument, as found in the writings of Smart, Reichenbach, Salmon, and Sellars, by arguing that such an unlimited demand for explanation leads to a demand for hidden variables, which runs contrary to at least one major school of thought in twentieth-century physics. I do not think that even these philosophers themselves wish to saddle realism with logical links to such consequences: but realist yearnings were born among the mistaken ideals of traditional metaphysics.

In his book *Between Science and Philosophy*, Smart gives two main arguments for realism. One is that only realism can respect the important distinction between *correct* and *merely useful* theories. He calls 'instrumentalist' any view that locates the importance of theories in their use, which requires only empirical adequacy, and not truth. But how can the instrumentalist explain the usefulness of his theories?

Consider a man (in the sixteenth century) who is a realist about the Copernican hypothesis but instrumentalist about the Ptolemaic one. He can explain the instrumental usefulness of the Ptolemaic system of epicycles because he can prove that the Ptolemaic system can produce almost the same predictions about the apparent motions of the planets as does the Copernican hypothesis. Hence the assumption of the realist truth of the Copernican hypothesis explains the instrumental usefulness of the Ptolemaic one. Such an explanation

of the instrumental usefulness of certain theories would not be possible if *all* theories were regarded as merely instrumental.¹⁴

What exactly is meant by 'such an explanation' in the last sentence? If no theory is assumed to be true, then no theory has its usefulness explained as following from the truth of another one—granted. But would we have less of an explanation of the usefulness of the Ptolemaic hypothesis if we began instead with the premiss that the Copernican gives implicitly a very accurate description of the motions of the planets as observed from earth? This would not assume the truth of Copernicus's heliocentric hypothesis, but would still entail that Ptolemy's simpler description was also a close approximation of those motions.

However, Smart would no doubt retort that such a response pushes the question only one step back: what explains the accuracy of predictions based on Copernicus's theory? If I say, the empirical adequacy of that theory, I have merely given a verbal explanation. For of course Smart does not mean to limit his question to actual predictions—it really concerns all actual and possible predictions and retrodictions. To put it quite concretely: what explains the fact that all observable planetary phenomena fit Copernicus's theory (if they do)? From the medieval debates, we recall the nominalist response that the basic regularities are merely brute regularities, and have no explanation. So here the anti-realist must similarly say: that the observable phenomena exhibit these regularities, because of which they fit the theory, is merely a brute fact, and may or may not have an explanation in terms of unobservable facts 'behind the phenomena'—it really does not matter to the goodness of the theory, nor to our understanding of the world.

Smart's main line of argument is addressed to exactly this point. In the same chapter he argues as follows. Suppose that we have a theory *T* which postulates micro-structure directly, and macro-structure indirectly. The statistical and approximate laws about macroscopic phenomena are only partially spelt out perhaps, and in any case derive from the precise (deterministic or statistical) laws about the basic entities. We now consider theory *T'*, which is part of *T*, and says only what *T* says about the macroscopic phenomena. (How *T'* should be characterized I shall leave open, for that does not affect the argument here.) Then he continues:

I would suggest that the realist could (say) . . . that the success of *T'* is explained by the fact that the original theory *T* is true of the things that it is ostensibly about; in other words by the fact that there really are electrons or whatever is postulated by the theory *T*. If there were no such things, and if *T* were not true in a realist way, would not the success of *T'* be quite inexplicable? One would have to suppose that there were innumerable lucky accidents about the behaviour mentioned in the observational vocabulary, so

that they behaved miraculously *as if* they were brought about by the non-existent things ostensibly talked about in the theoretical vocabulary.¹⁵

In other passages, Smart speaks similarly of 'cosmic coincidences'. The regularities in the observable phenomena must be explained in terms of deeper structure, for otherwise we are left with a belief in lucky accidents and coincidences on a cosmic scale.

I submit that if the demand for explanation implicit in these passages were precisely formulated, it would at once lead to absurdity. For if the mere fact of postulating regularities, without explanation, makes *T'* a poor theory, *T* will do no better. If, on the other hand, there is some precise limitation on what sorts of regularities can be postulated as basic, the context of the argument provides no reason to think that *T'* must automatically fare worse than *T*.

In any case, it seems to me that it is illegitimate to equate being a lucky accident, or a coincidence, with having no explanation. It was by coincidence that I met my friend in the market—but I can explain why I was there, and he can explain why he came, so together we can explain how this meeting happened. We call it a coincidence, not because the occurrence was inexplicable, but because we did not severally go to the market in order to meet.¹⁶ There cannot be a requirement upon science to provide a theoretical elimination of coincidences, or accidental correlations in general, for that does not even make sense. There is nothing here to motivate the demand for explanation, only a restatement in persuasive terms. . . .*

6 | Limits to Explanation: A Thought Experiment

Wilfrid Sellars was one of the leaders of the return to realism in philosophy of science and has, in his writings of the past three decades, developed a systematic and coherent scientific realism. I have discussed a number of his views and arguments elsewhere; but will here concentrate on some aspects that are closely related to the arguments of Smart, Reichenbach, and Salmon just examined.¹⁷ Let me begin by setting the stage in the way Sellars does.

There is a certain over-simplified picture of science, the 'levels picture', which pervades positivist writings and which Sellars successfully demolished.¹⁸ In that picture, singular observable facts ('this crow is black')

* We have omitted the following section in which van Fraassen criticizes Reichenbach's principle of the common cause and efforts by Salmon and others to argue that, because the principle is correct, we must postulate unobservable events and processes in order to explain correlations that would otherwise remain inexplicable.

are scientifically explained by general observable regularities ('all crows are black') which in turn are explained by highly theoretical hypotheses not restricted in what they say to the observable. The three levels are commonly called those of *fact*, of *empirical law*, and of *theory*. But, as Sellars points out, theories do not explain, or even entail such empirical laws—they only show why observable things obey these so-called laws to the extent they do.¹⁹ Indeed, perhaps we have no such empirical laws at all: all crows are black—except albinos; water boils at 100°C—provided atmospheric pressure is normal; a falling body accelerates—provided it is not intercepted, or attached to an aeroplane by a static line; and so forth. On the level of the observable we are liable to find only putative laws heavily subject to unwritten *ceteris paribus* qualifications.

This is, so far, only a methodological point. We do not really expect theories to 'save' our common everyday generalizations, for we ourselves have no confidence in their strict universality. But a theory which says that the micro-structure of things is subject to *some* exact, universal regularities, must imply the same for those things themselves. This, at least, is my reaction to the points so far. Sellars, however, sees an inherent inferiority in the description of the observable alone, an incompleteness which requires (*sub specie* the aims of science) an introduction of an unobservable reality behind the phenomena. This is brought out by an interesting 'thought-experiment'.

Imagine that at some early stage of chemistry it had been found that different samples of gold dissolve in *aqua regia* at different rates, although 'as far as can be observationally determined, the specimens and circumstances are identical'.²⁰ Imagine further that the response of chemistry to this problem was to postulate two distinct micro-structures for the different samples of gold. Observationally unpredictable variation in the rate of dissolution is explained by saying that the samples are mixtures (not compounds) of these two (observationally identical) substances, each of which has a fixed rate of dissolution.

In this case we have explanation through laws which have no observational counterparts that can play the same role. Indeed, no explanation seems possible unless we agree to find our physical variables outside the observable. But science aims to explain, must try to explain, and so must require a belief in this unobservable micro-structure. So Sellars contends.

There are at least three questions before us. Did this postulation of micro-structure really have no new consequences for the observable phenomena? Is there really such a demand upon science that it must explain—even if the means of explanation bring no gain in empirical predictions? And thirdly, could a *different* rationale exist for the use of a micro-structure picture in the development of a scientific theory in a case like this?

First, it seems to me that these hypothetical chemists did postulate new observable regularities as well. Suppose the two substances are A and

B, with dissolving rates x and $x + y$ and that every gold sample is a mixture of these substances. Then it follows that every gold sample dissolves at a rate no lower than x and no higher than $x + y$; and that between these two any value may be found—to within the limits of accuracy of gold mixing. None of this is implied by the data that different samples of gold have dissolved at various rates between x and $x + y$. So Sellars's first contention is false.

We may assume, for the sake of Sellars's example, that there is still no way of predicting dissolving rates any further. Is there then a categorical demand upon science to explain this variation which does not depend on other observable factors? . . . A precise version of such a demand (Reichenbach's principle of the common cause) could result automatically in a demand for hidden variables, providing a 'classical' underpinning for indeterministic theories. Sellars recognized very well that a demand for hidden variables would run counter to the main opinions current in quantum physics. Accordingly he mentions ' . . . the familiar point that the irreducibly and lawfully statistical ensembles of quantum-mechanical theory are mathematically inconsistent with the assumption of hidden variables.'²¹ Thus, he restricts the demand for explanation, in effect, to just those cases where it is *consistent* to add hidden variables to the theory. And consistency is surely a logical stopping-point.

This restriction unfortunately does not prevent the disaster. For while there are a number of proofs that hidden variables cannot be supplied so as to turn quantum mechanics into a classical sort of deterministic theory, those proofs are based on requirements much stronger than consistency. To give an example, one such assumption is that two distinct physical variables cannot have the same statistical distributions in measurement on all possible states.²² Thus it is assumed that, if we cannot point to some possible difference in empirical predictions, then there is no real difference at all. If such requirements were lifted, and consistency alone were the criterion, hidden variables could indeed be introduced. I think we must conclude that science, in contrast to scientific realism, does not place an overriding value on explanation in the absence of any gain for empirical results.

Thirdly, then, let us consider how an anti-realist could make sense of those hypothetical chemists' procedure. After pointing to the new empirical implications which I mentioned two paragraphs ago, he would point to methodological reasons. By imagining a certain sort of micro-structure for gold and other metals, say, we might arrive at a theory governing many observationally disparate substances; and this might then have implications for new, wider empirical regularities when such substances interact. This would only be a hope, of course; no hypothesis is guaranteed to be fruitful—but the point is that the true demand on science is not for explanation *as such*, but for imaginative pictures which have a hope of suggesting new statements of observable regularities and of correcting old

ones. This point is exactly the same as that for the principle of the common cause.

7 | Demons and the Ultimate Argument

Hilary Putnam, in the course of his discussions of realism in logic and mathematics, advanced several arguments for scientific realism as well. . . .

In . . . 'What is Mathematical Truth', Putnam gives . . . what I shall call the *Ultimate Argument*. He begins with a formulation of realism which he says he learned from Michael Dummett:

A realist (with respect to a given theory or discourse) holds that (1) the sentences of that theory are true or false; and (2) that what makes them true or false is something external—that is to say, it is not (in general) our sense data, actual or potential, or the structure of our minds, or our language, etc.²³

This formulation is quite different from the one I have given even if we instantiate it to the case in which that theory or discourse is science or scientific discourse. Because the wide discussion of Dummett's views has given some currency to his usage of these terms, and because Putnam begins his discussion in this way, we need to look carefully at this formulation.

In my view, Dummett's usage is quite idiosyncratic. Putnam's statement, though very brief, is essentially accurate. In his 'Realism', Dummett begins by describing various sorts of realism in the traditional fashion, as disputes over whether there really exist entities of a particular type. But he says that in some cases he wishes to discuss, such as the reality of the past and intuitionism in mathematics, the central issues seem to him to be about other questions. For this reason he proposes a new usage: he will take such disputes

as relating, not to a class of entities or a class of terms, but to a class of *statements*. . . . Realism I characterize as the belief that statements of the disputed class possess an objective truth-value, independently of our means of knowing it: they are true or false in virtue of a reality existing independently of us. The anti-realist opposes to this the view that statements of the disputed class are to be understood only by reference to the sort of thing which we count as evidence for a statement of that class.²⁴

Dummett himself notes at once that nominalists are realists in this sense.²⁵ If, for example, you say that abstract entities do not exist, and sets are abstract entities, hence sets do not exist, then you will certainly accord a

truth-value to all statements of set theory. It might be objected that if you take this position then you have a decision procedure for determining the truth-values of these statements (*false* for existentially quantified ones, *true* for universal ones, apply truth tables for the rest). Does that not mean that, on your view, the truth-values are not independent of our knowledge? Not at all; for you clearly believe that if we had not existed, and *a fortiori* had had no knowledge, the state of affairs with respect to abstract entities would be the same.

Has Dummett perhaps only laid down a necessary condition for realism, in his definition, for the sake of generality? I do not think so. In discussions of quantum mechanics we come across the view that the particles of microphysics are real, and obey the principles of the theory, but at any time *t* when 'particle *x* has exact momentum *p*' is true then 'particle *x* has position *q*' is neither true nor false. In any traditional sense, this is a realist position with respect to quantum mechanics.

We note also that Dummett has, at least in this passage, taken no care to exclude non-literal construals of the theory, as long as they are truth-valued. The two are not the same; when Strawson construed 'The king of France in 1905 is bald' as neither true nor false, he was not giving a non-literal construal of our language. On the other hand, people tend to fall back on non-literal construals typically in order to be able to say, 'properly construed, the theory is true.'

Perhaps Dummett is right in his assertion that what is really at stake, in realist disputes of various sorts, is questions about language—or, if not really at stake, at least the only serious philosophical problems in those neighbourhoods. Certainly the arguments in which he engages are profound, serious, and worthy of our attention. But it seems to me that his terminology ill accords with the traditional one. Certainly I wish to define scientific realism so that it need not imply that all statements in the theoretical language are true or false (only that they are all capable of being true or false, that is, there are conditions for each under which it has a truth-value); to imply nevertheless that the aim is that the theories should be true. And the contrary position of constructive empiricism is not anti-realist in Dummett's sense, since it also assumes scientific statements to have truth-conditions entirely independent of human activity or knowledge. But then, I do not conceive the dispute as being about language at all.

In any case Putnam himself does not stick with this weak formulation of Dummett's. A little later in the paper he directs himself to scientific realism *per se*, and formulates it in terms borrowed, he says, from Richard Boyd. The new formulation comes in the course of a new argument for scientific realism, which I shall call the Ultimate Argument:

the positive argument for realism is that it is the only philosophy that doesn't make the success of science a miracle. That terms in mature scientific the-

ories typically refer (this formulation is due to Richard Boyd), that the theories accepted in a mature science are typically approximately true, that the same term can refer to the same thing even when it occurs in different theories—these statements are viewed by the scientific realist not as necessary truths but as part of the only scientific explanation of the success of science, and hence as part of any adequate scientific description of science and its relations to its objects.²⁶

Science, apparently, is required to explain its own success. There is this regularity in the world, that scientific predictions are regularly fulfilled; and this regularity, too, needs an explanation. Once *that* is supplied we may perhaps hope to have reached the *terminus de jure*?

The explanation provided is a very traditional one—*adequatio ad rem*, the ‘adequacy’ of the theory to its objects, a kind of mirroring of the structure of things by the structure of ideas—Aquinas would have felt quite at home with it.

Well, let us accept for now this demand for a scientific explanation of the success of science. Let us also resist construing it as merely a re-statement of Smart’s ‘cosmic coincidence’ argument, and view it instead as the question why we have successful scientific theories at all. Will this realist explanation with the Scholastic look be a scientifically acceptable answer? I would like to point out that science is a biological phenomenon, an activity by one kind of organism which facilitates its interaction with the environment. And this makes me think that a very different kind of scientific explanation is required.

I can best make the point by contrasting two accounts of the mouse who runs from its enemy, the cat. St. Augustine already remarked on this phenomenon, and provided an intentional explanation: the mouse *perceives that* the cat is its enemy, hence the mouse runs. What is postulated here is the ‘adequacy’ of the mouse’s thought to the order of nature: the relation of enmity is correctly reflected in his mind. But the Darwinist says: Do not ask why the *mouse* runs from its enemy. Species which did not cope with their natural enemies no longer exist. That is why there are only ones who do.

In just the same way, I claim that the success of current scientific theories is no miracle. It is not even surprising to the scientific (Darwinist) mind. For any scientific theory is born into a life of fierce competition, a jungle red in tooth and claw. Only the successful theories survive—the ones which *in fact* latched on to actual regularities in nature.²⁷

■ | Notes

1. *Science, Perception and Reality* (New York: Humanities Press, 1962); cf. the footnote on p. 97. See also my review of his *Studies in Philosophy and its History*, in *Annals of Science*, January 1977.
2. Brian Ellis, *Rational Belief Systems* (Oxford: Blackwell, 1979), p. 28.
3. Hilary Putnam, *Mathematics, Matter and Method* (Cambridge: Cambridge University Press, 1975), Vol. 1, pp. 69f.
4. Putnam, *op. cit.*, p. 73, n. 29. The argument is reportedly developed at greater length in Boyd's forthcoming book *Realism and Scientific Epistemology* (Cambridge University Press).
5. Hartry Field has suggested that 'acceptance of a scientific theory involves the belief that it is true' be replaced by 'any reason to think that any part of a theory is not, or might not be, true, is reason not to accept it.' The drawback of this alternative is that it leaves open what epistemic attitude acceptance of a theory does involve. This question must also be answered, and as long as we are talking about full acceptance—as opposed to tentative or partial or otherwise qualified acceptance—I cannot see how a realist could do other than equate that attitude with full belief. (That theories believed to be false are used for practical problems, for example, classical mechanics for orbiting satellites, is of course a commonplace.) For if the aim is truth, and acceptance requires belief that the aim is served . . . I should also mention the statement of realism at the beginning of Richard Boyd, 'Realism, Underdetermination, and a Causal Theory of Evidence', *Noûs*, 7 (1973), 1–12. Except for some doubts about his use of the terms *explanation* and *causal relation* I intend my statement of realism to be entirely in accordance with his. Finally, see C. A. Hooker, 'Systematic Realism', *Synthese*, 26 (1974), 409–97; esp. pp. 409 and 426.
6. More typical of realism, it seems to me, is the sort of epistemology found in Clark Glymour's book, *Theory and Evidence* (Princeton: Princeton University Press, 1980), except of course that there it is fully and carefully developed in one specific fashion. (See esp. his chapter 'Why I Am Not a Bayesian' [reprinted in chapter 5 of this volume] for the present issue.) But I see no reason why a realist, as such, could not be a Bayesian of the type of Richard Jeffrey, even if the Bayesian position has in the past been linked with anti-realist and even instrumentalist views in philosophy of science.
7. G. Maxwell, 'The Ontological Status of Theoretical Entities', *Minnesota Studies in Philosophy of Science*, III (1962), p. 7. [An excerpt from Maxwell's paper is the first reading in this chapter.]
8. There is a great deal of recent work on the logic of vague predicates; especially important, to my mind, is that of Kit Fine ('Vagueness, Truth, and Logic', *Synthese*, 30 (1975), 265–300) and Hans Kamp. The latter is currently working on a new theory of vagueness that does justice to the 'vagueness of vagueness' and the context-dependence of standards of applicability for predicates.

9. Op. cit., p. 15. . . . At this point . . . I may be suspected of relying on modal distinctions which I criticize elsewhere. After all, I am making a distinction between human limitations and accidental factors. A certain apple was dropped into the sea in a bag of refuse, which sank; relative to that information it is necessary that no one ever observed the apple's core. That information, however, concerns an accident of history, and so it is not human limitations that rule out observation of the apple core. But unless I assert that some facts about humans are essential, or physically necessary, and others accidental, how can I make sense of this distinction? This question raises the difficulty of a philosophical retrenchment for modal language. This I believe to be possible through an ascent to pragmatics. In the present case, the answer would be, to speak very roughly, that the scientific theories we accept are a determining factor for the set of features of the human organism counted among the limitations to which we refer in using the term 'observable'. . . .

10. See n. 1 above.

11. Cf. P. Thagard, doctoral dissertation, University of Toronto, 1977, and 'The Best Explanation: Criteria for Theory Choice', *Journal of Philosophy*, 75 (1978), 76–92.

12. 'The Inference to the Best Explanation', *Philosophical Review*, 74 (1965), 88–95 and 'Knowledge, Inference, and Explanation', *American Philosophical Quarterly*, 5 (1968), 164–73. Harman's views were further developed in subsequent publications (Noûs, 1967; *Journal of Philosophy*, 1968; in M. Swain (ed.), *Induction*, 1970; in H.-N. Castañeda (ed.), *Action, Thought, and Reality*, 1975; and in his book *Thought*, Ch. 10). I shall not consider these further developments here.

13. See n. 11 above.

14. J. J. C. Smart, *Between Science and Philosophy* (New York: Random House, 1968), p. 151.

15. Ibid., pp. 150f.

16. This point is clearly made by Aristotle, *Physics*, II, Chs. 4–6 (see esp. 196^a 1–20; 196^b 20–197^a 12).

17. See my 'Wilfrid Sellars on Scientific Realism', *Dialogue*, 14 (1975), 606–16; W. Sellars, 'Is Scientific Realism Tenable?', pp. 307–34 in F. Suppe and P. Asquith (eds.), *PSA 1976* (East Lansing, Mich.: Philosophy of Science Association, 1977), vol. II, 307–34; and my 'On the Radical Incompleteness of the Manifest Image', *ibid.*, 335–43; and see n. 1 above.

18. W. Sellars, 'The Language of Theories', in his *Science, Perception, and Reality* (London: Routledge and Kegan Paul, 1963).

19. Op. cit., p. 121.

20. Ibid., p. 121.

21. Ibid., p. 123.

22. See my 'Semantic Analysis of Quantum Logic', in C. A. Hooker (ed.), *Contemporary Research in the Foundations and Philosophy of Quantum Theory* (Dordrecht: Reidel, 1973), Part III, Sects. 5 and 6.

23. See n. 3 above.

24. Michael Dummett, *Truth and Other Enigmas* (Cambridge, Mass.: Harvard University Press, 1978), p. 146 (see also pp. 358–61).

25. Dummett adds to the cited passage that he realizes that his characterization does not include all the disputes he had mentioned, and specifically excepts nominalism about abstract entities. However, he includes scientific realism as an example (op. cit., pp. 146f.).

26. See n. 4 above.

27. Of course, we can ask specifically why the *mouse* is one of the surviving species, how it survives, and answer this, on the basis of whatever scientific theory we accept, in terms of its brain and environment. The analogous question for theories would be why, say, Balmer's formula for the line spectrum of hydrogen survives as a successful hypothesis. In that case too we explain, on the basis of the physics we accept now, why the spacing of those lines satisfies the formula. Both the question and the answer are very different from the global question of the success of science, and the global answer of realism. The realist may now make the *further* objection that the anti-realist cannot answer the question about the mouse specifically, nor the one about Balmer's formula, in this fashion, since the answer is in part an assertion that the scientific theory, used as basis of the explanation, is true. This is a quite different argument, which I . . . take up in Ch. 4, Sect. 4, and Ch. 5 [of *The Scientific Image*].

In his most recent publications and lectures Hilary Putnam has drawn a distinction between two doctrines, metaphysical realism and internal realism. He denies the former and identifies his preceding scientific realism as the latter. While I have at present no commitment to either side of the metaphysical dispute, I am very much in sympathy with the critique of Platonism in philosophy of mathematics, which forms part of Putnam's arguments. Our disagreement about scientific (internal) realism would remain, of course, whenever we came down to earth after deciding to agree or disagree about metaphysical realism, or even about whether this distinction makes sense at all.

Realism versus Constructive Empiricism

The demise of logical positivism has been followed by a rising tide of scientific realism. Bas van Fraassen is to be congratulated for swimming against that tide. But we must also ask whether he manages to make much headway. I shall argue that he does not. My first section explores van Fraassen's rather attenuated antirealism and the distinction between truth and empirical adequacy on which it depends. My second section argues that van Fraassen succeeds no better than his predecessors in answering a major objection to antirealism. My third section examines the link between realism and explanation and van Fraassen's attempt to sever that link.

I | Truth, Empirical Adequacy, Empirical Equivalence

Scientific realism is an old issue, and over the years both realism and antirealism have taken various forms. Van Fraassen defines realism thus: "*Science aims to give us, in its theories, a literally true story of what the world is like; and acceptance of a scientific theory involves the belief that it is true.*" He says that this is a minimal formulation which "can be agreed to by anyone who considers himself a scientific realist."¹ [1066]* Later, however, van Fraassen extends this minimal formulation by adding to it a realist 'demand for explanation.' As we will see, his antirealism stems in large part from criticisms of this demand. As we will also see, his version of the demand is an absurdly strong one.

What is the nature of van Fraassen's antirealism? The most radical opponents of realism (the instrumentalists) deny that scientific theories

FROM Paul M. Churchland and Clifford A. Hooker, eds., *Images of Science* (Chicago: University of Chicago Press, 1985), 197–221.

* The numbers in brackets refer to pages in this volume.

have truth-values at all. Van Fraassen's antirealism is not of this radical kind. He accepts a "literal construal of the language of science" whereby "the apparent statements of science really are statements, *capable of* being true or false" (p. 10) [1068]. In the same vein, he rejects positivist interpretations of scientific language, whereby the 'real meaning' of theoretical assertions is somehow cashed out in terms of the observable:

Most specifically, if a theory says that something exists, then a literal construal may elaborate on what that something is, but will not remove the implication of existence. . . . If the theory's statements include "There are electrons," then the theory says that there are electrons. If in addition they include "Electrons are not planets," then the theory says, in part, that there are entities other than planets. (p. 11) [1068–69]

Thus, contrary to the positivists, two theories may say exactly the same thing about the observable yet remain distinct and perhaps incompatible theories.

All this puts van Fraassen firmly in the realist camp as far as the *interpretation* of scientific theories is concerned.² His antirealism proceeds entirely on the *epistemological or methodological* level. (The same can be said of the antirealism espoused by Larry Laudan in *Progress and Its Problems*.) He thinks that, although scientific theories are capable of literal truth, they "need not be true to be good" (p. 10) [1067–68]. Accordingly, it is not the aim of science to provide true theories, and to accept a theory is not to believe it to be true. What matters in science is that theories are correct so far as the observations and experiments go. Hence, constructive empiricism: "*Science aims to give us theories which are empirically adequate; and acceptance of a theory involves as belief only that it is empirically adequate*" (p. 12) [1069]. A theory is empirically adequate "exactly if what it says about the observable things and events in this world, is true—exactly if it 'saves the phenomena' " (p. 12) [1069].

The distinction between truth and empirical adequacy, and hence between realism and constructive empiricism, is a subtle one. For theories about the observable, truth and empirical adequacy coincide (p. 21) [1077]. For theories about the unobservable, truth entails empirical adequacy but not vice versa: such a theory may be empirically adequate yet false. Accordingly, to believe that a theory about the unobservable is true is more risky than to believe that it is empirically adequate. Not that the latter is without risk: empirical adequacy "goes far beyond what we can know at any given time" since it requires that the theory save all the phenomena in its field, past, present, and future, not merely all actually observed phenomena (p. 69). Now, the chief difficulty for realism has always been skeptical arguments to the effect that we can never know a scientific theory to be true nor ever be rationally warranted in accepting, however tentatively, a theory as true. This is as much a difficulty for con-

structive empiricism. The same skeptical arguments might be used to show that we can never know a scientific theory to be empirically adequate nor ever be rationally warranted in accepting, however tentatively, a theory as empirically adequate. Van Fraassen insists, however, that the positions are different:

There does remain the fact that . . . in accepting any theory as empirically adequate, I am sticking my neck out. There is no argument there for belief in the truth of the accepted theories, since it is not an epistemological principle that one might as well hang for a sheep as for a lamb. (p. 72)

Epistemological or not, the principle that one might as well hang for a sheep as for a lamb is a pretty sensible one. Given two criminal acts A and B whose risks of detection and subsequent penalties are the same, but where A yields a greater gain than B, the sensible criminal will do A. But are the risks and penalties of realism and constructive empiricism the same? And does realism bring with it gains that constructive empiricism does not? Van Fraassen addresses these questions; to evaluate the cogency of his position we must address them too.

Suppose the realist tentatively accepts a theory as true, while the constructive empiricist tentatively accepts it as empirically adequate. The realist does take a greater risk. But he takes no greater risk of being detected in error *on empirical grounds*. So, given strict empiricism (the principle that only evidence should determine theory choice), it seems that we might as well be hung for the realist sheep as for the constructive empiricist lamb.

The trouble is, van Fraassen argues, that realism and strict empiricism do not mix and that realism must pay the penalty of rejecting strict empiricism. He makes the point by considering the case of empirically equivalent yet incompatible theories. This is not, of course, the humdrum case where the *available* evidence fails to discriminate between two incompatible theories. This case need not trouble the realist, who may always hope to show that the two theories are not empirically equivalent and then press for an experimental decision between them. Rather, it is the esoteric case where such hopes are unfounded, where two incompatible theories say exactly the same things about *all* matters observational. The constructive empiricist could accept both theories (believe both to be empirically adequate); the realist cannot on pain of contradiction believe both to be true. But how is the realist to choose between them? In the nature of the case, empirical evidence cannot guide his choice, which must therefore be made on nonevidential grounds. Realism runs counter to strict empiricism and allows nonevidential or 'metaphysical' considerations to intrude into matters of theory choice.

How might the realist respond to this? Presumably, as a realist, he will have no truck with the positivist idea that empirically equivalent the-

ories are really the same theory and not incompatible after all. Nor, as a realist, will he have any truck with the related idea (perhaps it is the same idea in new dress) that there are no 'verification-transcendent truth-conditions' and therefore no truth of the matter for the two theories to disagree about. These ideas, after all, seem to entail that Berkeley's immaterialism is really the same theory as the commonsense belief in independently existing material objects or that there is no truth of the matter for Berkeley and commonsense to disagree about. And these conclusions are anathema to the commonsense realist, let alone the scientific one.³

Taking a cue from this example, one might wonder whether the problem is philosophical or metaphysical rather than scientific, in which case metaphysical considerations would not be an intrusion after all. How often have empirically equivalent but incompatible theories occurred in real science? Van Fraassen gives one example, and it is a notorious one. Newton hypothesized that the center of gravity of the solar system is at rest in absolute space. He also pointed out that the appearances would be no different if that center were moving through absolute space at any constant velocity v . So all of the theories $TN(v)$ —Newton's theories of mechanics and gravitation plus the postulate that the center of gravity of the solar system has constant absolute velocity v for any v —were claimed by Newton to be empirically equivalent (p. 46).

Van Fraassen's account can be disputed: Newton only claimed the empirical equivalence of the theories $TN(v)$ *so far as appearances within the solar system are concerned*. Hypothesize that some other star is at rest in absolute space, for example, and the empirical equivalence vanishes: if the solar system has any nonzero velocity, then it will approach or recede from that star and, given sufficient time, the effects of this will become apparent.

Here I resort to a realist ploy whose efficacy van Fraassen considers. This is to say that, where equivalent theories occur, by *extending* these theories (that is, embedding them in wider theories) their equivalence will disappear (that is, the wider theories will not be empirically equivalent). In the trivial example just cited, the wider theories are formed simply by adding the statement that some star is at absolute rest to each of the existing theories. The example is trivial because we can, by the Newtonian principle of relativity, construct empirically equivalent theories to each of these wider theories (including the only empirically adequate one): simply consider theories attributing an absolute velocity v to the center of mass of the extended system consisting of the solar system and the star. The process can be continued (assuming the number of masses is finite) until they are all taken into account. And then, again by the Newtonian principle of relativity, we will have an infinite family of empirically equivalent theories each of which consists of Newton's laws plus the hypothesis that the center of mass of *the universe* has velocity v for any value of v .

Van Fraassen considers a more interesting extension or embedding of

his Newtonian example, the attempt to combine it with Maxwell's electromagnetism where forces depend upon velocities and not upon accelerations as in Newton. This feature made it possible to devise experiments to detect absolute velocities. The null results of such experiments were an important factor in leading scientists to abandon the Newtonian doctrines of absolute space and time in favor of relativistic ones. And this was to abandon *all* of the empirically equivalent Newtonian theories. Van Fraassen asks us to imagine, however, that null results had not been obtained, that, on the contrary, an absolute velocity for the center of mass of the solar system had been measured. Here it might seem that one of the empirically equivalent Newtonian theories had been confirmed and the rest refuted, and hence that they were not empirically equivalent after all. Van Fraassen finds this reasoning spurious (p. 49). But I find his reasoning, if not spurious, at least hard to follow. Operating within this piece of science fiction (or, rather, history-of-science fiction), he says that we could make compensating adjustments in electromagnetic theory so as to retain whichever of the empirically equivalent Newtonian theories we like. In other words, had the history of science been different, we could construct a new family of empirically equivalent combinations of mechanics and electromagnetism.

But, first, van Fraassen has done nothing to impugn the fact that his empirically equivalent Newtonian theories, when combined with Maxwell's electromagnetism, ceased to be empirically equivalent. Second, could the Newtonian readily have accepted that electromagnetic forces depend upon absolute velocities rather than absolute accelerations? Van Fraassen concedes that, had his piece of history-of-science fiction occurred, it would have "upset even Newton's deepest convictions about the relativity of motion" (p. 48). But did not these convictions follow from Newton's laws of mechanics and the doctrine of absolute space? Last, and perhaps most important, all of this is a piece of history-of-science fiction: the historical facts are that in this notorious real example of empirical equivalence, the only good example known to me, the actual development of science removed the problem.

Van Fraassen has a further retort to the idea that empirical equivalence can be removed by extension or embedding. He can say that it is the empirical adequacy of the extended theories which counts and that one should accept the victor only as empirically adequate, never as true. And, to offset the scarcity of empirically equivalent theories in real science, he can point to the fact that we can artificially concoct empirically equivalent alternatives to any theory by resorting to notorious logical tricks. The simplest such trick is to conjoin any theory with "The Absolute is lazy" to form a new theory empirically equivalent with the original.

The standard response to such tricks is to eliminate the concocted theory on the ground of its reduced simplicity or unity. Van Fraassen does not object to the appeal to simplicity, but he insists that simplicity is a

pragmatic virtue of a theory which has nothing to do with that theory's truth or likelihood of being true. The realist, for whom accepting a theory is believing it true, must forge a link between simplicity and truth if he is to appeal to the former. And the link can be forged only by a metaphysical principle:

Simplicity . . . is obviously a criterion in theory choice, or at least a term in theory appraisal. For that reason, some . . . suggest that simple theories are more likely to be true. But it is surely absurd to think that the world is more likely to be simple than complicated (unless one has certain metaphysical or theological views not usually accepted as legitimate factors in scientific inference). The point is that the virtue, or patchwork of virtues, indicated by the term is a factor in theory appraisal, but does not . . . make a theory more likely to be true (or empirically adequate). (p. 90)

So the argument seems to be this: the realist can solve the problem of empirical equivalence only by appealing to simplicity; but he can appeal to simplicity only if he assumes a metaphysical principle ("Nature is simple" or some such); realism therefore involves an illegitimate intrusion of metaphysics into science and the abandonment of strict empiricism.

Is the constructive empiricist in any better position? Presumably he, too, will prefer a respectable theory to an artificially concocted empirically equivalent alternative. He, too, will appeal to simplicity and abandon strict empiricism. But he, apparently, can do this in good conscience, cheerfully admitting that pragmatic virtues such as simplicity have nothing to do with the real aim of science, empirical adequacy. Indeed, how could simplicity have anything to do with that aim? To say that the simpler of two empirically equivalent theories is more likely to be empirically adequate is to contradict oneself.

Returning to the realist, there are several ways he might respond to van Fraassen's argument. The first is simply to admit that there is nothing to choose between empirically equivalent theories. This is hardly satisfactory in view of the ubiquity of the logical tricks. The second is to spice scientific realism with a dash of pragmatism, admitting that there is nothing to choose on realist grounds between empirically equivalent theories but preferring some on the pragmatic ground of simplicity. Despite van Fraassen's argument, I see no reason why the realist cannot appeal to pragmatic virtues just as the constructive empiricist does. The third response is to say that simplicity is not, after all, a merely pragmatic virtue. Realists and constructive empiricists alike value empirical strength; they value it for different reasons, but both connect it with the central aim of science. Is it not a sufficient reason to eliminate concocted alternatives to existing theories that they are not empirically stronger than the theories from which they are concocted?

The realist has a problem here, however. Insofar as simplicity and

strength go together (and they do not always), simplicity is not merely a pragmatic virtue. But insofar as simplicity and strength go together, simplicity and truth cannot: the stronger theory is, in some intuitive sense at least, less likely to be true. And here lies the problem for any realist seeking to forge a link between simplicity and truth. Yet the problem may not be completely intractable. "Nature is simple" is a metaphysical principle and a hopelessly vague one to boot. But scientists have made various attempts to say more precisely what it means and to construct theories which conform to it.⁴ This transforms it into a metaphysical principle which can, at first remove so to speak, be empirically assessed: roughly speaking, it is acceptable metaphysics if theories constructed under its aegis are empirically successful, while theories which violate it are not.⁵ In our postpositivistic age, we should not regard the intrusion of this kind of metaphysical principle into science as illegitimate. If vague appeals to simplicity can be transformed into precise principles of theory construction and if such principles are acceptable (in the sense roughly defined), then the virtue they indicate is not merely pragmatic. It may not be absurd to think that Nature is simple (in some carefully specified sense or senses), if we can point to the empirical success of science in vindication of our belief.

I do not know whether this third response, which I have merely sketched, will work in the end. Perhaps it could be shown (though it would be a far from trivial task) that, for any precise and acceptable sense of the term *simple*, one could concoct empirically equivalent and equally simple theories. I would not see this as the demise of scientific realism, for (and this is the second response again) I cannot see why the realist is barred from invoking a pragmatic virtue to deal with the problem of empirical equivalence just as the constructive empiricist does.

II | Theory and Observation

Antirealists need to draw a dichotomy between theory and observation. Van Fraassen is no exception: after all, he could not even distinguish truth from empirical adequacy without it. An old and powerful objection to antirealist views is that no such dichotomy exists. How does van Fraassen deal with this objection?

He first agrees that no such dichotomy can be drawn in scientific language, agreeing with the realist that "All our language is thoroughly theory-infected" (p. 14) [1071] and pointing out against the positivist that highly theoretical assertions can be made using only so-called 'observational vocabulary' (pp. 54–55). (Here I was reminded of how Popper formulated "There exists an omnipotent, omnipresent, and omniscient personal spirit" in a physicalistic observation language.)⁶

Van Fraassen does insist, however, that some objects and/or events

are observable and some not. He concedes the familiar realist point that there is a continuous spectrum between 'directly observing' an object and 'indirectly detecting' it using instruments. This only shows that *observable* is a vague predicate. But a vague predicate is perfectly usable provided it has clear cases and clear noncases—and this one has:

A look through a telescope at the moons of Jupiter seems to me a clear case of observation, since astronauts will no doubt be able to see them as well from close up. But the purported observation of micro-particles in a cloud chamber seems to me a clearly different case—if our theory about what happens there is right. . . . while the particle is detected by means of the cloud chamber, and the detection is based on observation, it is clearly not a case of the particle's being observed. (pp. 16–17) [1073]

What if we had microscopic or electron-microscopic eyes? (Actually, we do, only they are not built into our heads!) Could we not then observe things which at present we can only detect, showing that they were not unobservable in principle? Lockean speculations like this merely change the subject:

The human organism is, from the point of view of physics, a certain kind of measuring apparatus. As such it has certain inherent limitations—which will be described in detail in the final physics and biology. It is these limitations to which the "able" in "observable" refers—our limitations, *qua* human beings. (p. 17) [1074]

But current physics and biology tell us that what is observable by humans varies (some of us are color-blind) and depends on our particular evolutionary history (other organisms can observe things we cannot). So, even if we can draw a rough and species-specific distinction between what is observable by humans and what is not, should any philosophical significance be attached to it?

Van Fraassen agrees with the realists against the idealists that it has no *ontological* significance: things that humans do not happen to be able to observe may nonetheless exist (p. 18) [1074]. (Actually, I will argue later, there are problems about van Fraassen's making this concession.) But he wants to give the distinction an *epistemological* significance: humans should never believe to be true a theory about what they cannot observe; instead, they should believe such theories only to be empirically adequate, to tell the truth about what they can observe (p. 18) [1074].

Can a distinction which is admitted to be rough-and-ready, species-specific, and of no ontological significance really bear such an epistemological burden? Van Fraassen gives an example of a so-called inference to the best explanation:

I hear scratching in the wall, the patter of little feet at midnight, my cheese disappears—and I infer that a mouse has come to live with me. Not merely that these apparent signs of mousely presence will continue, not merely that all the observable phenomena will be as if there is a mouse; but that there really is a mouse. (pp. 19–20) [1076]⁷

Will not the same style of argument lead us to the conclusion that there really are electrons (or whatever)? Van Fraassen thinks not. He accepts “inference to the best explanation” but puts his own gloss upon it: such inferences should (and do) only lead us to accept the best explanation as empirically adequate (p. 20) [1076]. If the best explanation is a theory about the observable, then empirical adequacy and truth coincide and we can (and do) conclude that there really is a mouse (or whatever). But if the best explanation is a theory about the unobservable, empirical adequacy and truth do not coincide and we cannot (and do not) conclude that there really are electrons (or whatever).

There is an empirical claim here (about what scientists actually do infer) and also a methodological claim (about what they ought to infer). I find the methodological claim quite unreasonable. On any plausible theory of evidential support, one would have to admit that there could be far better evidence for an explanation couched in terms of unobservables than for an explanation couched in terms of observables. Is the evidence for the existence of electrons better or worse than the evidence for the existence of the yeti or of the mouse in van Fraassen’s wainscoting? It is a curious sort of empiricism which sets aside the weight of *available* evidence on the ground that a casual observer might one day see his mouse or yeti, while the scientist can never see (but can only detect) his electrons.

Van Fraassen’s factual claim (that scientists do infer only the empirical adequacy of theories about the unobservable but never their truth) is even harder to swallow. Admittedly, I have not done a sociological survey to settle the matter. And, even if such a survey were to reveal, as I believe it would, that realism is the instinctive philosophy of working scientists, this would not of course settle the methodological question. But to indicate how difficult it is to avoid realist ways of thinking and talking, let us see how van Fraassen thinks and talks. He talks of *detecting* an electron in a cloud chamber. Can one say truly that one has detected an object without also believing it to be true that the object really exists? Later he describes how Millikan *measured* the charge of the electron (pp. 75–77). Did not Millikan think it true, and does not anyone who accepts Millikan’s results think it true, that electrons exist and carry a certain charge? Can one say truly that one has measured some feature of an object without also believing that the object really exists?

I shall quote at some length what I *think* is van Fraassen’s answer to very obvious questions like these:

The working scientist is totally immersed in the scientific world-picture. And not only he—to varying degrees, so are we all. . . . But immersion in the theoretical world-picture does not preclude “bracketing” its ontological implications. . . . To someone immersed in that world-picture, the distinction between *electron* and *flying horse* is as clear as between *racehorse* and *flying horse*: the first corresponds to something in the actual world, and the other does not. While immersed in the theory, and addressing oneself solely to problems in the domain of the theory, this objectivity of *electron* is not and cannot be qualified. *But this is so whether or not one is committed to the truth of the theory.* It is so not only for someone who believes, full stop, that the theory is true, but also for . . . someone who . . . holds commitment to the truth of the theory in abeyance. For to say that someone is immersed in theory . . . is not to describe his epistemic commitment. . . . it is possible even after total immersion in the world of science . . . to limit one’s epistemic commitment while remaining a functioning member of the scientific community. (pp. 80–83)

This is, I fear, nothing but a sleight-of-hand and an endorsement of philosophical schizophrenia. The sleight-of-hand converts belief in the reality of electrons (belief in the objectivity of *electron*, belief that the term *electron* corresponds to something in the actual world) into belief in, belief full stop in, and finally commitment to something called “the theory of electrons.” But there have been several theories about electrons, and no scientist believes them all to be true. As for the most up-to-date theory about electrons, sensible scientists would do well not to believe it to be wholly true either, for details of it are quite likely to be further refined. All this is quite consistent with a pretty firm belief in the reality of electrons, with a refusal to “bracket” this particular ontological implication of science. The philosophical schizophrenia stems from talk of immersion (even total immersion) in the “scientific world-picture” or the “world of science.” These metaphors are meant to suggest, if I understand them rightly, that scientists should believe in electrons or whatever while immersed in their scientific work, but should become agnostic about everything they cannot observe once they leave their laboratories. I suppose that split-minded scientists like this are possible, but I wonder whether they are desirable.

Finally in this section, I want to argue that van Fraassen’s treatment of the observable/unobservable distinction verges on the incoherent. He insists that what is observable by humans is a “function of facts about us *qua* organisms in the world,” so that it is for *science* to tell us what is observable and what is not (pp. 57–58).⁸ Now, suppose some theory *T* does distinguish “the observable which it postulates from the whole it postulates” (p. 59). *T* might even be van Fraassen’s “final physics and biology,” if such a theory is possible. *T* will say, among other things, that

A is observable by humans, while B is not. Of course, if we are to use *T* to delineate the observable, we must *accept* it. But van Fraassen cannot have us accept it as true, since it concerns in part the unobservable *B*. The constructive empiricist can accept *T* only as empirically adequate, that is, believe to be true only what *T* says about the observable. But “B is not observable by humans” cannot, on pain of contradiction, be a statement about something observable by humans. And, in general, the consistent constructive empiricist cannot believe it to be true that *anything* is unobservable by humans. And, if this is so, the consistent constructive empiricist cannot draw a workable observable/unobservable dichotomy at all.

It might be objected that

1 B is not observable by humans

is logically equivalent with

2 Everything observable by humans is distinct from B

[and] since (2) is a statement about the observable, so is the logically equivalent (1). But even accepting that there is a sense in which (2) is “about” the observable, it is *also* about the unobservable *B* and therefore cannot be accepted as true by the constructive empiricist.

Nor does it help if we say that “observable by humans” is an “observational predicate,” that we humans can tell from observation that a thing is observable by us. For one thing, this marks a retreat from van Fraassen’s insistence that there is no observable/unobservable dichotomy in scientific *language*. For another thing, “observable by humans” will be a nonstandard observational predicate whose negation is not also observational, a predicate akin, for example, to the predicate “is an inscription of finite length.” For we cannot observe that *anything* has a property without also observing that thing. Anyone who claims to have *observed* that something is unobservable contradicts himself. But if “unobservable by humans” is *not* an observational predicate, our conclusion stands. We can grant that “observable by humans” is an observational predicate so that the constructive empiricist can accept as true, on the basis of observation, statements of the form “A is observable by humans.” But the consistent constructive empiricist cannot accept as true, on the basis of observation or anything else, a statement of the form “B is not observable by humans.” Constructive empiricism requires a dichotomy which it cannot consistently draw.

III | Realism and Explanation

Realism and explanation are doubly linked. Realists think science explains facts about the world, and they think realist philosophy of science explains facts about science. I will consider the latter claim first.

The claim is that only a realist philosophy of science can explain the fact that science has had a great deal of predictive success. If the unobservables postulated by (successful) science really exist and if what (successful) science says about them is true or nearly so, then we can explain predictive success. Otherwise, such success is just a lucky accident. As Putnam famously remarks, realism is “the only philosophy that doesn’t make the success of science a miracle” (cited on p. 39) [1083].

Van Fraassen gives short shrift to this Ultimate Argument for realism:

The explanation provided is a very traditional one—*adequatio ad rem*, the “adequacy” of the theory to its objects, a kind of mirroring of the structure of things by the structure of ideas—Aquinas would have felt quite at home with it.

. . . Will this realist explanation with the Scholastic look be a scientifically acceptable answer? I would like to point out that science is a biological phenomenon, an activity by one kind of organism which facilitates its interaction with the environment. And this makes me think that a very different kind of scientific explanation is required.

I can best make the point by contrasting two accounts of the mouse who runs from its enemy, the cat. St. Augustine . . . provided an intentional explanation: the mouse *perceives* that the cat is an enemy, hence the mouse runs. What is postulated here is the “adequacy” of the mouse’s thought to the order of nature: the relation of enmity is correctly reflected in his mind. But the Darwinist says: Do not ask why the *mouse* runs from its enemy. Species which did not cope with their natural enemies no longer exist. That is why there are only ones who do.

In just the same way, I claim that the success of current scientific theories is no miracle. It is not even surprising to the scientific (Darwinist) mind. For any scientific theory is born into a life of fierce competition, a jungle red in tooth and claw. Only the successful theories survive—the ones which *in fact* latched on to actual regularities in nature. (pp. 39–40) [1084]

Amusing though this is, it does no more than play cat-and-mouse with the argument. The scientist does ask why the mouse runs from the cat and answers in roughly the terms made fun of here: the mouse perceives the cat, perceives the cat as an enemy, and runs. This does not commit the

scientist to ascribing thoughts, adequate or otherwise, to the mouse: his response might be quite instinctive. But with this proviso, there is nothing unscientific or un-Darwinian about this kind of explanation. Of course, the Darwinian question is not "Why does the mouse run away from the cat?" but, rather, "How did this piece of mouse behavior evolve?" The Darwinian answers this question roughly in the terms suggested by van Fraassen: given an environment full of mouse-hunting cats, cat-fleeing mice are more likely to survive, reproduce, and pass their cat-fleeing behavior on to future generations. But the Darwinian explanation is not a substitute for the "intensional" one, for they are addressed to quite different questions. The Darwinian explains what the "intensionalist" postulates: that the mouse's perceiving the cat as an enemy (or, better, the mouse's genetically programmed behavioral response to cats) is adequate to the order of nature.

Just as with cats and mice, so also with scientific success. It is one thing to explain why some theory is successful and quite another to explain why only successful theories survive. Van Fraassen's Darwinian explanation of the latter can be accepted by realist and antirealist alike. But to say that only successful theories are allowed to survive is not to explain why any particular theory is successful.

Not that a realist explanation of this in terms of the theory's *adequatio ad rem* will do as it stands. The Ultimate Argument is actually very old, and a brief look at an old example of it should give us pause. Christopher Clavius (in his *Commentary on the Sphere of Sacrobosco* of 1581) said it was incredible to suppose that Ptolemaic astronomy could correctly predict eclipses even though its eccentrics and epicycles were mere figments. But the eccentrics and epicycles were figments, and it was no miracle at all that a geometrical model expressly devised to yield some phenomenal regularity (periodic eclipses) should be successful in doing so. It is different, however, if a theory devised to accommodate some phenomenal regularities should turn out to predict *new* regularities. The realist has a ready explanation: the entities postulated by the theory really exist, and what the theory says about them is true (or nearly so). The antirealist seems forced to say that figments dreamed up for one purpose have turned out, miraculously, to be well adapted for a quite different purpose. So it was that thoughtful realists such as William Whewell distinguished two kinds of predictive success (predicting known effects and predicting new ones) and argued that the antirealist cannot explain the latter. So it was that a thoughtful antirealist such as Duhem, seeing the force of the argument, came to spice his instrumentalism with a whiff of realism: a theory is able to make successful *novel* predictions because it is not "a purely artificial system" but, rather, "a natural classification [whose] principles express profound and real relations among things."⁹

As this brief historical excursion shows, the only form of the Ultimate Argument which *might* work is that which focuses on *novel* predictive

success. Yet this focus is lacking in recent discussions, both from prominent defenders of the argument (such as Putnam) and from prominent critics of it (such as Laudan).¹⁰ Difficulties remain, of course, not least that of making precise the intuitive distinction between known effects and novel predictions. These difficulties notwithstanding, van Fraassen has done nothing to impugn the Ultimate Argument in its refined form.

The Ultimate Argument proceeds on the metalevel: epistemology is to be naturalized, and philosophy of science is to explain facts about science. But there is a more direct argument, which proceeds from the assumption that science should explain facts about the world. The connection between the demand for explanation and the realist demand for true theories is apparently very obvious. An explanation is not adequate unless what does the explaining is true. So, given that theories figure in explanations, adequate explanations require true theories. It is worth noting that Duhem found the argument cogent and confessed that since in his view science did not aim at true theories, it could not explain anything either. Others adopt the curious view that science does aim truly to describe the world but cannot really explain any features of it. Van Fraassen's response to our simple argument is twofold. He attacks the realist demand for explanation; and he argues that explanation, where it can be had, does not require true theories, that explanatory power is a pragmatic virtue for which an empirically adequate theory will do just as well as a true one.

He softens us up with a linguistic point. We still speak of explanations even when we think the explanatory principles false:

I say that Newton could explain the tides, that he had an explanation of the tides, that he did explain the tides. In the same breath I can add that this theory is, after all, not correct. Hence I would be inconsistent if by the former I meant that Newton had a true theory which explained the tides. (p. 99)

Quite so. We can speak without contradiction of a false explanation because truth is not a *defining* condition of explanation but an *adequacy* condition upon it. (In a similar way, we can say that Bode's Law is false without contradicting ourselves.)* Van Fraassen says that we may agree

* Despite its name, Bode's law is not a law. It is a formula that says that the distances of the planets from the sun are in the following ratio: $(0 + 4)$, $(3 + 4)$, $(6 + 4)$, $(12 + 4)$, $(24 + 4)$, and so on. The formula fits the planets from Mercury to Uranus quite well, if we include the asteroids as a "missing planet" between Mars and Jupiter. But Neptune violates the series, since its distance calculated from Bode's law is 388, and its actual distance is about 300. So the formula is false and hence not a law. The interesting question, of course, is whether we would count Bode's law as a genuine law if there were no exceptions to it. Our reluctance to do so would stem largely from the fact that, as yet, no one has given a convincing explanation of why it holds (in the cases that it does). It stands isolated from the rest of science as a cosmic coincidence. See chapter 7 for further discussion of the distinction between laws and accidental generalizations.

that a theory is false without at all undermining our previous assertion that it explained many phenomena (p. 98). But could we say that a theory *adequately* explained many phenomena even though it is false? Realists think not. Scientists appear to agree: no modern astronomical text cites the vortex theory of planetary motion as the explanation of why the planets all go around the sun in the same direction, though some may have no other explanation of this fact. If this is wrong, and truth is not required for adequate explanation, then it will take more than a linguistic point to show it—van Fraassen gives us more.

First, he attacks the realist demand for explanation. He claims that in science this demand is severely limited, that explanation is not a “pre-eminent” or “rock-bottom” scientific virtue:

If explanation of the facts were required in the way consistency with the facts is, then every theory would have to explain every fact in its domain. Newton would have had to add an explanation of gravity to his celestial mechanics before presenting it at all. (p. 94)

But Newton did present his theory. He “declined to explain,” admitting famously that he had “not been able to discover the cause of . . . gravity” (p. 94). Thus:

Newton’s theory of gravitation . . . did not (in the opinion of Newton or his contemporaries) contain an explanation of gravitational phenomena, but only a description. (p. 112)

It is the same in modern physics, where the “unlimited demand for explanation leads to a demand for hidden variables, which runs counter to at least one major school of thought” (p. 23) [1077]. And, quite generally, to demand that regularities be explained and shown to be more than cosmic coincidences is self-defeating. For what of the regularities postulated to do the explaining?

Something is obviously wrong somewhere in all this. On the one hand, Newton explained the tides (p. 99) and, on the other hand, Newton’s theory did not explain gravitational phenomena at all (p. 112). What has gone wrong is a tacit conflation of realism with essentialism, of the demand for explanation with the demand for ultimate explanation. This will take a bit of explaining.¹¹

Suppose we explain the phenomenal regularity that sticks look bent when half-immersed in water by postulating, among other things, that unobservable light rays refract when passing through media of different densities. This is not, of course, to explain the refraction of light, though we might then try to do so. But at any point in our explanatory endeavors there will be things for which we have no explanations, namely the deepest explanatory principles we have reached at that point. One realist response

to this situation is to demand that these principles require explanation in their turn. Another realist response, quite the antithesis of the first, is to demand that our deepest explanatory principles should require no explanation, that they should somehow be ultimate or self-explanatory. This second response is central to the tradition of Aristotelian essentialism, the tradition which holds that the only genuine explanations are ultimate or self-explanatory explanations.

The essentialist tradition is buttressed by rhetorical questions like these. Do we *really* explain why sticks look bent in water by postulating some mysterious and unexplained law of refraction? Did Newton *really* explain the tides by postulating his mysterious and unexplained force of gravity? The intuition is that a *real* explanation does not remove one mystery by postulating another one. And behind this intuition lies another one: the intuition that a real explanation should serve the *pragmatic* function of removing puzzlement, setting the curiosity of the inquirer at rest. It is because nonultimate scientific explanations do not serve this pragmatic function, do not remove puzzlement so much as relocate and enhance it, that they are said not to be real explanations at all.

I use the term *pragmatic* here advisedly and roughly as van Fraassen uses it (sometimes). Whether an explanation removes puzzlement depends very much on the person we are considering. What relieves one man's puzzlement may enhance the next woman's. I dare say that some of the most efficacious puzzlement relievers in the history of thought have been explanations which are not scientific at all and which are, from a scientific point of view, quite inadequate (what about "God moves in mysterious ways," said in explanation of anything whatever which is puzzling?). I dare say that on occasions the incurious have had their puzzlement removed by a scientific explanation—but they should not have. For if it is feelings of puzzlement we want to get rid of, we should turn not to science but to the whiskey bottle!

I think that *essentialism* and the intuitions which lie behind it are to be rejected. And I think it one of Newton's chief claims to methodological fame that he was among the first to see this. Newton admitted that he could not explain gravity *and* that gravity was a perfectly proper thing to try to explain (since it was not an "essential property" of matter). Yet in the same breath he insisted that his theory of gravity did explain celestial motions and the tides:

Hitherto we have *explained* the phenomena of the heavens and of our sea by the power of gravity, but we have not yet assigned the cause of this power . . . hitherto I have not been able to discover the cause of . . . gravity from phenomena, and I frame no hypotheses . . . to us it is enough that gravity does really exist, and act according to the laws we have explained, and abundantly serves to *account* for all the motions of the celestial bodies, and of our sea.¹²

Elsewhere Newton contrasted his procedure with that of his opponents, the Cartesian essentialists: he gives precise deductive explanations; they mutter about essences and can explain nothing:

To tell us that every Species of Things is endow'd with an occult specifick Quality by which it acts and produces manifest Effects, is to tell us nothing: But to derive two or three general Principles of Motion from Phaenomena, and afterwards to tell us how the Properties and Actions of all corporeal Things follow from these manifest Principles, would be a very great step in Philosophy, though the Causes of those Principles were not yet discover'd: And therefore I scruple not to propose the Principles of Motion above mention'd, they being of very general Extent, and leave their Causes to be found out.¹⁵

Van Fraassen quotes from the first of these famous passages (p. 94), but he misunderstands Newton (as Duhem and others have misunderstood him) in a way that can only stem from a tacit conflation of realism with essentialism. He says that Newton "decline[d] to explain" (p. 94) and that Newton's opinion was that his theory "did not contain an explanation of gravitational phenomena, but only a description" (p. 112). But Newton explicitly claimed to have explained or accounted for gravitational phenomena such as the tides *by* describing precisely how gravity works. The antithesis between explanation and description is quite illusory: we explain one thing *by* describing another. Newton did decline to explain gravity. But to take this as a confession that nothing can be explained by the law of gravity is to father upon Newton a view that was not his.

It is, however, the view which van Fraassen calls the "realist demand for explanation" and attacks. He formulates the demand innocently enough as "every theory should explain every fact in its domain" and then takes "every fact" to include *the theory itself* (p. 94, quoted earlier). Only a theory which was somehow self-explanatory could meet this demand. But one can demand explanation without also demanding ultimate or self-explanatory explanation, as Newton tried to teach us. Van Fraassen's rejection of the latter demand leaves the former quite intact.

The tacit conflation of explanation with ultimate explanation also emerges from the delightful joke which forms the last chapter of van Fraassen's book. There, taking his cue from the remark that "everyone believed in the existence of God until the Boyle lectures proved it" (p. 229), he modifies Aquinas's Five Ways into proofs of scientific realism. What gets proved is actually Aristotelian essentialism, which is perhaps not surprising considering the provenance of the arguments. The First Way gives the flavor of the whole:

So I argue: Everything that is to be explained, is to be explained by something else. That some things are to be explained is evident, for the regularities in

natural phenomena are obvious to the senses and surprising to the intellect. So we must either proceed to infinity, or arrive at something which explains, but is not itself, a regularity in the natural phenomena. However, in this we cannot proceed to infinity. (pp. 205–6)

Now, what gets “proved” here is that science can achieve something which explains but is not itself a regularity in the natural phenomena, something, in other words, which is not itself to be explained. What gets proved is the essentialist view that science can achieve ultimate explanation. But the “proof” has a missing premise (required to obtain the statement beginning with “So”): the essentialist principle that A does not really explain B if A also requires explanation and has not received it. Van Fraassen hopes that readers skeptical of ultimate explanations will reject the conclusion, miss the missing premise, and infer by *modus tollens* that science cannot explain anything at all. But I am spoiling a clever joke.

I cannot leave this topic without saying a word about van Fraassen’s favorite example of how science has transcended realist demands for explanation, quantum mechanics. Hidden variable explanations of quantum mechanics are said to “run counter to at least one modern school of thought” (p. 23) [1077]. A philosophy of science is not refuted by pointing out that it runs counter to a scientific school of thought, not even a dominant school of thought. But this school can point to *proofs* that hidden variable theories are impossible, and this should give the realist pause. Van Fraassen’s comments on these proofs (which he seems to approve of) reveal some interesting things. One proof apparently assumes that “if we cannot point to some possible differences in empirical predictions, then there is no real difference at all” between two theories (p. 34) [1081]. In endorsing this proof, van Fraassen’s earlier resolute antipositivism has wavered, for this assumption requires a positivist reinterpretation of scientific language to show that empirically equivalent theories are really the same theory. Not that realists would be too happy with an explanation of quantum mechanics which was demonstrably empirically equivalent with it: such an explanation could have no *independent* evidence in its favor. Now, if we assume the truth and completeness of quantum mechanics (or even its empirical adequacy and completeness), we will be able to prove that no explanation of it could have independent evidence in its favor. And if we assume that quantum mechanics is not only true and complete but also ultimately so, we will be able to prove that no explanation of it (independently confirmable or not) will be adequate. These various assumptions are no part of quantum mechanics; rather, they are philosophical assertions about quantum mechanics. Insofar as the various proofs rest upon assumptions like these (I do not know whether they do), they are not so much proofs as philosophical arguments, and pretty questionable ones to boot. Finally, the issue of determinism, important though it is in other contexts, is something of a red herring in this context. It is true that

some hidden-variable theorists wanted a deterministic explanation of quantum mechanics. But there is no *a priori* reason why a deeper explanation of quantum mechanics has to be deterministic.

Van Fraassen says nothing to impugn a modest realist demand for nonultimate explanation. And realists can defend such a demand by pointing to cases where the attempt to explain, even to explain theories regarded as empirically adequate, has paid off handsomely. Van Fraassen is not impressed with the argument:

Paid off handsomely, how? Paid off in new theories we have more reason to believe empirically adequate. But in that case even the anti-realist, when asked questions about *methodology* will *ex cathedra* counsel the search for explanation! We might even suggest a loyalty oath for scientists, if realism is so efficacious. (p. 93)

Realists might retort that explanation has a payoff in terms of *understanding* the world—but that is unlikely to impress van Fraassen. And realists who are also empiricists are impaled on the horns of a dilemma here (as the case of the hidden-variable theories suggested). Realists who are also empiricists will want any proposed explanation to yield empirical regularities other than those it was devised to explain; otherwise there could be no independent evidence for the truth of the explanation. If an explanation does yield them, then the constructive empiricist can value it, too, but for its predictive rather than its explanatory power. If an explanation does not yield them, then it should be rejected as mere “metaphysical baggage.” Heads constructive empiricism wins, tails realism loses:

I think we must conclude that science, in contrast to scientific realism, does not place an overriding value on explanation in the absence of any gain for empirical results. . . . the point is that the true demand on science is not for explanation *as such*, but for imaginative pictures which have a hope of suggesting new statements of observable regularities and of correcting old ones. (p. 34) [1081–82]

This true demand is not, it seems, to be vouchsafed to scientists themselves. They are to take an oath of loyalty to realism, the desire to understand the world, and the search for explanatory truths. Realism is the constructive empiricist’s Noble Lie, propounded *ex cathedra* in case scientists should find the true aim (enhancing the empirical adequacy of “imaginative pictures”) uninspiring! More seriously, is van Fraassen right to say that “the interpretation of science, and the correct view of its methodology, are two separate topics” (p. 93)? I think it preferable to have an interpretation of science which harmonizes with methodological pronouncements.

At any rate, this is part of what van Fraassen means when he calls

explanation a *pragmatic* virtue. The search for explanation *works* because theories which are good explainers will *ipso facto* be good savers of phenomena and the real “name of the game is saving the phenomena” (p. 93). But this is not all that he means. The rest is meant to undercut the simple realist idea that adequate explanations must contain theories that are true (or nearly so). Van Fraassen defends a “pragmatic” analysis of explanation according to which theories do not figure in explanations at all but somehow lie behind or underpin them. And he says that good explanations can be underpinned by empirically adequate theories just as well as true ones. At least, I think that is what is going on.

Charles Morris divided the study of language into syntax, semantics, and pragmatics. The last was meant to deal, among other things, with *context dependence*, as when the truth or falsity of “I’m hungry” depends upon the context of utterance, upon who says it and when. In philosophical circles, “pragmatic” (or better “merely pragmatic”) has also come to mean “useful but not true.” There is no obvious connection between these two philosophical usages. Utterances can express both truths and falsehoods in virtue of contextual factors. And an utterance can be useful but not true whether or not contextual factors enter into it. (“He went that-away” may be useful for diverting the pursuer though it is false; it is also heavily context-dependent. “John Brown took the road to California” may be similarly useful but false, and it is less context-dependent.)

Van Fraassen thinks that explanation is also heavily context-dependent. He begins from Bromberger’s puzzle about explanatory asymmetries: the height of the flagpole explains the length of its shadow but not vice versa, though the two deductions may be structurally identical. (Actually, van Fraassen also tells a blue story which is meant to show—I do not think it does show it—that there are “contexts” in which the length of the shadow does explain the height of the flagpole.) The obvious solution to this puzzle is an appeal to causality: explanations exhibit causes, while nonexplanatory deductions do not. But this is to jump out of the contextual frying pan into the contextual fire, for which factor is picked out as “the cause” also varies enormously with the context:

... [T]he salient feature picked out as “the cause” ... is salient to a given person because of his orientation, his interests, and various other peculiarities in the way he ... comes to know the problem—contextual factors. (p. 125)

Now, John Stuart Mill, who first drew attention to this kind of thing, insisted that only the entire constellation of factors, amounting to a sufficient condition for the event to be explained, is really entitled to be called the cause of it. And John Mackie, with Mill in spirit though more amenable to ordinary ways of talking, said that a cause is an insufficient but necessary part of an unnecessary but sufficient condition. Both Mill and Mackie can admit that contextual factors may influence *which event a*

person wants explained. And this is enough to dispose of many of the usual examples adduced to demonstrate context dependence of the explanations given (car crashes, fires, and such like).¹⁴

Van Fraassen takes a different course: he accepts the context dependence of explanations and tries to make it more precise. An explanation "is an *answer* . . . to a why-question" (p. 134). Every why-question has a *topic* (if we ask, "Why *P*?" the topic is *P*), an implied *contrast-class* (what we actually ask is, "Why *P* rather than *Q*, *R*, etc.?"), and an implied relation of *explanatory-relevance* which determines what shall count as a possible answer to the question (pp. 142–43). The topic, contrast class, and relation of explanatory relevance all depend upon the context, in particular upon "a certain body *K* of accepted background theory and factual information," which in turn depends "on who the questioner and his audience are" (p. 145). This gets complicated. The upshot of it is that

the discussion of explanation went wrong at the very beginning when explanation was conceived of as a relationship like description: a relation between theory and fact. Really it is a three-term relation, between theory, fact, and context. . . . So to say that a given theory can be used to explain a certain fact, is always elliptic for: there is a proposition which is a telling answer, relative to this theory, to the request for information about certain facts (those counted as relevant for *this* question) that bear on a comparison between this fact which is the case, and certain (contextually specified) alternatives which are not the case. (p. 156)

Is it *really* elliptic for all this? At the outset of his discussion, in order to combat "the increasing sense of unreality" the usual examples bring, van Fraassen sets forth three "workaday examples of scientific explanation" (pp. 101–3). And he manages to set these examples forth *without mentioning contextual factors at all*. There are why-questions all right, but there are no implied contrast classes, relevance relations, or anything else which depends "on who the questioner and his audience are." Contextual complications have little to do with explanation in science, if van Fraassen's own "workaday examples" are anything to go by.

Is it true to say that explanation was ever "conceived of as a relationship like description: a relation between theory and fact"? The orthodox account says that explanations are arguments in which three things figure: theories or general laws, initial conditions specifying the cause of the event being explained, and in the conclusion a statement of the event being explained. Explanations *contain* descriptions, but they are not *like* them.

Van Fraassen (his earlier examples notwithstanding) wants to drop the theories out of explanations and relegate them to the "context," to the "background information" relative to which why-questions are asked and answered. (Here I was reminded of the Wittgensteinians, who assimilate theories to rules of inference and insist that these rules do not figure as

premises in the inferences constructed in accordance with them.)¹⁵ Still, one might think, the theories must be true if the explanations proffered in the light of them (or constructed in accordance with them) are to be correct. Van Fraassen thinks not. Empirically adequate and empirically strong theories will do just as well as true ones. His argument seems to be as follows. The fact to be explained is always an observable fact. The facts cited in explanation of it are also always observable facts. So what the theory has to get right, to be a good explainer, are just the observable facts. And empirically adequate theories, by definition, do just that:

So scientific explanation is not (pure) science but an application of science. It is a use of science to satisfy certain of our desires; and these desires are quite specific in a specific context, but they are always desires for *descriptive information*. . . . in each case, a success of explanation is a success of *adequate and informative description* [of the phenomena]. And while it is true that we seek for explanation, the value of this search for science is that the search for explanation is *ipso facto* a search for empirically adequate, empirically strong theories. (pp. 156–57)¹⁶

I am not sure that I have got the argument right here—but, if I have, there is lots wrong with it. Sometimes we explain observable facts by citing other observable facts (and laws). But this is not always the case, though it tends to be the case in the usual philosopher's examples, which bring an "increasing sense of unreality" to the subject. The flagpole is a good example again: both its height and the length of its shadow are presumably observable facts. (I can scarcely bring myself to mention the "explanation" of why some bird is black, which consists in pointing out that it's a raven and they are all black!) But, in van Fraassen's own examples of scientific explanations, there are initial conditions such as: the specific heats of water and copper are 1 and 0.1, respectively; the earth's magnetic field at a certain point has a vertical component of approximately $5/10^5$ Tesla; the energy levels associated with stable electron orbits in hydrogen atoms take the form $E_n = -E_0/n^2$ where E_0 is called the ground state energy. These initial conditions are *generalized* ones because the facts to be explained are *general* ones (with the possible exception of the second). But, setting aside the problem of their generality, they do not look much like observable facts, and their provision does not look much like providing descriptions of the observable phenomena. Nor, in science, is it always observable phenomena that we try to explain: we sometimes try to explain theories.

Van Fraassen not only thinks that explanation is a pragmatic affair in Morris's sense (context dependence), he also thinks that explanatory power is one of the *pragmatic virtues*, concerning which he says in general:

In so far as they go beyond consistency, empirical adequacy, and empirical strength, they do not concern the relation between the theory and the world,

but rather the use and usefulness of the theory; they provide reasons to prefer the theory independently of questions of truth. (p. 88)

If what I have said about scientific explanation is right, then the explanatory power of a scientific theory does depend on whether it tells the truth about the unobservable and therefore does go beyond empirical adequacy and empirical strength. But van Fraassen is obviously right when he says:

Nor can there be any question of explanatory success as providing evidence for the truth of a theory that goes beyond any evidence we have for its providing an adequate description of the phenomena. (pp. 156–57)

This is obviously right, because empirical adequacy is *defined* as correctness so far as the observable evidence is concerned. Explaining things cannot provide a special sort of *evidence* that theories are true rather than just empirically adequate. Realists, made of sterner stuff than constructive empiricists, still demand that a theory be true for the explanations in which it figures to be adequate. And realism carries this much metaphysical baggage: realists can point to no evidence over and above evidence of empirical adequacy that their sterner requirement has been met.

But there is excess baggage of a different kind in the constructive empiricist position. There is, above all, the philosophical excess baggage of defending an observable/unobservable distinction and giving it crucial epistemological significance. There is the excess baggage of providing an alternative to the obvious realist explanation of science's novel predictive success. And there is the excess baggage of a complex account of the pragmatics of explanation.

I suggested earlier that, in comparing constructive empiricism with scientific realism, we should assess the risks, penalties, and gains associated with each. The risks have been discussed, as have the penalties in the form of philosophical "excess baggage" of various kinds. As to the respective gains (or losses), I can only repeat a hackneyed point. The realist values theoretical science as an attempt to *understand* the world and sees continuity between commonsense and scientific knowledge. The constructive empiricist, browbeaten as much by the positivist emphasis on prediction as by esoteric problems in interpreting quantum theory realistically, jettisons understanding and seeks to drive a wedge between theoretical science and commonsense (taking his excess baggage aboard to do so).

Let me conclude by agreeing with van Fraassen that it would be a pity if scientific realism were to become a philosophical dogma. Bas van Fraassen's book certainly roused me from any dogmatic slumber to which I might have been prone. It has, in fact, given me sleepless nights! His antirealism is more viable than earlier antirealist positions. But, in philosophy of science as well as in science, viability directly depends on weakness. Constructive empiricism is weaker than earlier antirealist views in

all kinds of ways, and correspondingly closer to realism. This is why I conclude, undogmatically I hope, that realism emerges a little bloodied but unbowed from its encounter with constructive empiricism.¹⁷

Notes

1. *The Scientific Image* (Oxford: Clarendon Press, 1980), 8 [1066]. Henceforth, all page numbers in the text refer to this book. [The numbers in brackets refer to the excerpt from this book included as the second reading in this chapter.]

2. Here and in what follows I ignore, through lack of space, a central feature of van Fraassen's position, his preference for a semantic approach to scientific theories whereby they emerge as sets of models rather than as sets of (true or false) sentences. I have two excuses for this. First, in much of his own discussion van Fraassen ignores it too, and talks as though theories consisted of true or false sentences. Second, and more important, I think that there is little to choose between the two approaches from a logical point of view. As van Fraassen himself once wrote, "There are natural interrelations between the two approaches: an axiomatic theory may be characterized by the class of interpretations which satisfy it, and an interpretation may be characterized by the set of sentences which it satisfies. . . . These interrelations make implausible any claim of superiority for either approach" ("On the Extension of Beth's Semantics of Physical Theories," *Philosophy of Science* 37 [Sept. 1970]: 325–39; cf. p. 326). I am indebted, both for the general point and for the reference to van Fraassen's endorsement of it, to John Worrall's review article of *The Scientific Image*, ["An Unreal Image,"] *British Journal for the Philosophy of Science* [35 (1984): 65–80].

3. Here I assume that Berkeley's immaterialism is empirically equivalent to commonsense realism. I am not sure that this is so. Immaterialism can be formulated so as to be consistent with all possible experience and so irrefutable by it. But empirical adequacy should require more than mere consistency with the evidence; it should require (at least) that the theory in question entail the evidence. It might be argued that commonsense realism entails consequences about the stability and reobservability of physical objects which Berkeley's immaterialism does not. Berkeley does invoke God's benevolence to "explain" *post factum* the stability of tables and trees. But Berkeley cannot predict it because of his admission that God might always make an exception to his "laws of nature" and work a miracle instead.

4. Einstein was always appealing to simplicity or unity. For an analysis of how these vague appeals were articulated into quite powerful principles of theory construction, see E. Zahar, "Why Did Einstein's Programme Supersede Lorentz's (II)?" *British Journal for the Philosophy of Science* 24 (1973): 223–62.

Philosophers of science still, of course, lack a precise and general account of what simplicity is; perhaps there is none to be had and simplicity is, as van Fraassen says, a patchwork of virtues, some pragmatic and some not. Popper's identification of simplicity and strength works nicely sometimes ("All swans are white" is simpler and stronger than "All non-Australasian swans are white") and badly at other times ("All swans are white and ferocious" is less simple and stronger than "All swans are white").

5. For more detail of how certain metaphysical principles can be rationally assessed in this way, see Watkins, "Confirmable and Influential Metaphysics," *Mind* 67 (July 1958): 344–65, especially pp. 363–65. Watkins does not apply these ideas to principles of simplicity.

6. See K. R. Popper, *Conjectures and Refutations* (London: Routledge and Kegan Paul, 1963), 274–76.

7. Here, incidentally, there is a curious prejudice in favor of vision. True, I have not *seen* the mouse, but have I not *heard* it, and is this not a way of observing it? There is a curious tension in the view that, though we can see (and touch) things, we never hear (or taste or smell) things but only the *noises* they make (or the tastes and smells they emit). Notoriously, one way to resolve the tension is to say that we never really see things, either, but only visions (visual sense-data) caused by them (and the same will have to go, even more implausibly, for touch). That way leads to idealism. Realists resolve the tension by saying that we can hear, taste, and smell things as well as see and touch them. Van Fraassen, for all his talk about hearing "an apparent sign of mousely presence" rather than a mouse, is once again with the realists. He says that "sense-data, I am sure, do not exist" (p. 72). And he has to be with the realists if truth and empirical adequacy are to be the same so far as observable things are concerned. If all observable phenomena were only apparent signs of mousely presence (mousely visions, mousely noises, mousely smells, etc.), then "All observable phenomena are as if there is a mouse in the wainscoting" would not entail "There is a mouse in the wainscoting," contrary to what van Fraassen says on p. 21 [1077].

8. In case anyone is reminded here of the talk of "observables" in quantum mechanics, we should remind ourselves that the so-called "observables" of quantum mechanics are in the present context remotely calculable theoretical quantities. If electrons are not observable, neither is their charge, momentum, or spin.

9. P. Duhem, *The Aim and Structure of Physical Theory* (Princeton, N.J.: Princeton University Press, 1954), 28 (see also pp. 297ff).

10. For example, Laudan presents as historical counterexamples successful theories which did not genuinely refer and were not true or nearly true ("A Confutation of Convergent Realism," *Philosophy of Science* 48, no. 1 [March 1981], 19–49). But few, arguably none, of the theories cited had any *novel* predictive success. Laudan also saddles the realist with the principle that successful reference alone breeds success. I do not know if any realist has thought this, but no realist should think it. For, as Laudan shows, we can construct referring theories which will be quite unsuccessful: take a successful theory containing the term *t* and negate it. Successful reference is a necessary condition for truth (or near truth) but not a sufficient one. And when it comes to success, it is the truth (or near truth) of what a theory says about its theoretical entities which counts, not whether those entities exist. [Laudan's paper follows next in this chapter.]

11. Further detail can be found in my "Explanation, Description and Scientific Realism," *Scientia* 112 (1977), 727–55.

12. *Principia*, Book 3, General Scholium; Motte's translation, revised by Cajori, vol. 2 (Berkeley and Los Angeles: University of California Press, 1962), 546–47. *Italics mine*. Also important is the famous passage where Newton says he will treat

forces "not physically but mathematically" (*Principia*, Book 1, Definition 8; Cajori, vol. 1, 5–6). Newton is saying that he will describe how gravity works in precise mathematical terms, rather than try to explain it physically. He is often misinterpreted as saying that gravity does not really exist.

13. *Opticks*, Book 3, Query 31; Dover edition (New York, 1952), 401–2. The most extended defense of Newton and attack on Cartesian essentialism is, of course, Roger Cotes's preface to the second edition of the *Principia*. Yet Cotes is sometimes interpreted as defending the essentialist view that gravity is, after all that Newton had said to the contrary, essential to matter (for example, by Popper in *Conjectures and Refutations*, 106). It is unlikely that Newton would have allowed Cotes to defend a view he himself had specifically denied. Cotes himself specifically denied that this was the view he was defending, in a letter to Samuel Clarke, who had questioned him on the point (see Cajori's appendix, n. 6; vol. 2, 634–35).

14. This is shown in John Worrall's review article, referred to in note 2 above. Worrall also shows that we can dispose in the same way of van Fraassen's story about the length of a flagpole's shadow explaining its height.

15. For a discussion of this view, see my "Wittgensteinian Instrumentalism," *Theoria* 46 (1980), pts. 2–3, 65–105.

16. The italics in this quotation are mine. And I felt justified in adding "of the phenomena" to "description" because in the preceding sentence (not quoted here) the phrase "adequate description of the phenomena" occurs.

17. Previously published in a shorter form as "Constructive Empiricism versus Scientific Realism," *Philosophical Quarterly* 32 (July 1982): 262–71. Reproduced by permission. I am grateful to Cliff Hooker for his helpful suggestions about what I might focus attention upon in the paper and to Greg Currie, Bob Durrant, and Martin Fricke for comments on earlier versions.

A Confutation of Convergent Realism

The positive argument for realism is that it is the only philosophy that doesn't make the success of science a miracle.

—H. Putnam (1975)

1 | The Problem

It is becoming increasingly common to suggest that epistemological realism is an empirical hypothesis, grounded in, and to be authenticated by its ability to explain the workings of science. A growing number of philosophers (including Boyd, Newton-Smith, Shimony, Putnam, Friedman and Niiniluoto) have argued that the theses of epistemic realism are open to empirical test. The suggestion that epistemological doctrines have much the same empirical status as the sciences is a welcome one: for, whether it stands up to detailed scrutiny or not, it marks a significant facing-up by the philosophical community to one of the most neglected (and most notorious) problems of philosophy: the status of epistemological claims.

But there are potential hazards as well as advantages associated with the 'scientizing' of epistemology. Specifically, once one concedes that epistemic doctrines are to be tested in the court of experience, it is possible that one's favorite epistemic theories may be refuted rather than confirmed. It is the thesis of this paper that precisely such a fate afflicts a form of realism advocated by those who have been in the vanguard of the move to show that realism is supported by an empirical study of the development of science. Specifically, I shall show that epistemic realism, at least in certain of its extant forms, is neither supported by, nor has it made sense of, much of the available historical evidence.

2 | Convergent Realism

Like other philosophical *-isms*, the term 'realism' covers a variety of sins. Many of these will not be at issue here. For instance, 'semantic realism' (in brief, the claim that all theories have truth values and that some

theories—we know not which—are true) is not in dispute. Nor shall I discuss what one might call 'intentional realism' (i.e., the view that theories are generally intended by their proponents to assert the existence of entities corresponding to the terms in those theories). What I shall focus on instead are certain forms of *epistemological* realism. As Hilary Putnam has pointed out, although such realism has become increasingly fashionable, "very little is said about what realism is" (1978). The lack of specificity about what realism asserts makes it difficult to evaluate its claims, since many formulations are too vague and sketchy to get a grip on. At the same time, any efforts to formulate the realist position with greater precision lay the critic open to charges of attacking a straw man. In the course of this paper, I shall attribute several theses to the realists. Although there is probably no realist who subscribes to all of them, most of them have been defended by some self-avowed realist or other; taken together, they are perhaps closest to that version of realism advocated by Putnam, Boyd and Newton-Smith. Although I believe the views I shall be discussing can be legitimately attributed to certain contemporary philosophers (and will frequently cite the textual evidence for such attributions), it is not crucial to my case that such attributions can be made. Nor will I claim to do justice to the complex epistemologies of those whose work I will criticize. My aim, rather, is to explore certain epistemic claims which those who are realists might be tempted (and in some cases have been tempted) to embrace. If my arguments are sound, we will discover that some of the most intuitively tempting versions of realism prove to be chimeras.

The form of realism I shall discuss involves variants of the following claims:

- R1 Scientific theories (at least in the 'mature' sciences) are typically approximately true and more recent theories are closer to the truth than older theories in the same domain.
- R2 The observational and theoretical terms within the theories of a mature science genuinely refer (roughly, there are substances in the world that correspond to the ontologies presumed by our best theories).
- R3 Successive theories in any mature science will be such that they 'preserve' the theoretical relations and the apparent referents of earlier theories (i.e., earlier theories will be 'limiting cases' of later theories).¹
- R4 Acceptable new theories do and should explain why their predecessors were successful insofar as they were successful.

To these semantic, methodological and epistemic theses is conjoined an important meta-philosophical claim about how realism is to be evaluated and assessed. Specifically, it is maintained that:

- R5 Theses (R1)–(R4) entail that ('mature') scientific theories should be successful; indeed, these theses constitute the best, if not the only, explanation for the success of science. The empirical success of science (in the sense of giving detailed explanations and accurate predictions) accordingly provides striking empirical confirmation for realism.

I shall call the position delineated by (R1) to (R5) *convergent epistemological realism*, or CER for short. Many recent proponents of CER maintain that (R1), (R2), (R3), and (R4) are empirical hypotheses which, via the linkages postulated in (R5), can be tested by an investigation of science itself. They propose two elaborate abductive [inference-to-the-best-explanation] arguments. The structure of the first, which is germane to (R1) and (R2), is something like this:

ARGUMENT I

- 1 If scientific theories are approximately true, they will typically be empirically successful.
 - 2 If the central terms in scientific theories genuinely refer, those theories will generally be empirically successful.
 - 3 Scientific theories are empirically successful.
-
- 4 (Probably) Theories are approximately true and their terms genuinely refer.

The argument relevant to (R3) is of slightly different form, specifically:

ARGUMENT II

- 1 If the earlier theories in a 'mature' science are approximately true and if the central terms of those theories genuinely refer, then later more successful theories in the same science will preserve the earlier theories as limiting cases.
 - 2 Scientists seek to preserve earlier theories as limiting cases and generally succeed.
-
- 3 (Probably) Earlier theories in a 'mature' science are approximately true and genuinely referential.

Taking the success of present and past theories as givens, proponents of CER claim that *if* CER were true, it would follow that the success and the progressive success of science would be a matter of course. Equally, they allege that if CER were false, the success of science would be 'miraculous' and without explanation.² Because (on their view) CER explains the fact that science is successful, the theses of CER are thereby confirmed by the success of science and non-realist epistemologies are discredited by

the latter's alleged inability to explain both the success of current theories and the progress which science historically exhibits.

As Putnam and certain others (e.g., Newton-Smith) see it, the fact that statements about reference (R2, R3) or about approximate truth (R1, R3) function in the explanation of a contingent state of affairs, establishes that "the notions of 'truth' and 'reference' have a causal explanatory role in epistemology" (Putnam 1978, p. 21).³ In one fell swoop, both epistemology and semantics are 'naturalized' and, to top it all off, we get an explanation of the success of science into the bargain!

The central question before us is whether the realist's assertions about the interrelations between truth, reference and success are sound. It will be the burden of this paper to raise doubts about both I and II. Specifically, I shall argue that *four* of the five premises of those abductions are either false or too ambiguous to be acceptable. I shall also seek to show that, even if the premises were true, they would not warrant the conclusions which realists draw from them. . . .

3 | Reference and Success

The specifically referential side of the 'empirical' argument for realism has been developed chiefly by Putnam, who talks explicitly of reference rather more than most realists. On the other hand, reference is usually implicitly smuggled in, since most realists subscribe to the (ultimately referential) thesis that "the world probably contains entities very like those postulated by our most successful theories."

If R2 is to fulfill Putnam's ambition that reference can explain the success of science, and that the success of science establishes the presumptive truth of R2, it seems he must subscribe to claims similar to these:

- S1 The theories in the advanced or mature sciences are successful.
- S2 A theory whose central terms genuinely refer will be a successful theory.
- S3 If a theory is successful, we can reasonably infer that its central terms genuinely refer.
- S4 All the central terms in theories in the mature sciences do refer.

There are complex interconnections here. (S2) and (S4) explain (S1), while (S1) and (S3) provide the warrant for (S4). Reference explains success and success warrants a presumption of reference. The arguments are plausible, given the premises. But there is the rub, for with the possible exception of (S1), none of the premises is acceptable.

The first and toughest nut to crack involves getting clearer about the nature of that 'success' which realists are concerned to explain. Although Putnam, Sellars and Boyd all take the success of certain sciences as a

given, they say little about what this success amounts to. So far as I can see, they are working with a largely *pragmatic* notion to be cashed out in terms of a theory's workability or applicability. On this account, we would say that a theory is successful if it makes substantially correct predictions, if it leads to efficacious interventions in the natural order, if it passes a battery of standard tests. One would like to be able to be more specific about what success amounts to, but the lack of a coherent theory of confirmation makes further specificity very difficult.

Moreover, the realist must be wary—at least for these purposes—of adopting too strict a notion of success, for a highly robust and stringent construal of 'success' would defeat the realist's purposes. What he wants to explain, after all, is why science in general has worked so well. If he were to adopt a very demanding characterization of success (such as those advocated by inductive logicians or Popperians) then it would probably turn out that science has been largely 'unsuccessful' (because it does not have high confirmation) and the realist's avowed explanandum would thus be a non-problem. Accordingly, I shall assume that a theory is 'successful' so long as it has worked well, i.e., so long as it has functioned in a variety of explanatory contexts, has led to confirmed predictions and has been of broad explanatory scope. As I understand the realist's position, his concern is to explain why certain theories have enjoyed this kind of success.

If we construe 'success' in this way, (S1) can be conceded. Whether one's criterion of success is broad explanatory scope, possession of a large number of confirming instances, or conferring manipulative or predictive control, it is clear that science is, by and large, a successful activity.

What about (S2)? I am not certain that any realist would or should endorse it, although it is a perfectly natural construal of the realist's claim that 'reference explains success'. The notion of reference that is involved here is highly complex and unsatisfactory in significant respects. Without endorsing it, I shall use it frequently in the ensuing discussion. The realist sense of reference is a rather liberal one, according to which the terms in a theory may be genuinely referring even if many of the claims the theory makes about the entities to which it refers are false. Provided that there are entities which "approximately fit" a theory's description of them, Putnam's charitable account of reference allows us to say that the terms of a theory genuinely refer.⁴ On this account (and these are Putnam's examples), Bohr's 'electron', Newton's 'mass', Mendel's 'gene', and Dalton's 'atom' are all referring terms, while 'phlogiston' and 'aether' are not (Putnam 1978, pp. 20–22).

Are genuinely referential theories (i.e., theories whose central terms genuinely refer) invariably or even generally successful at the empirical level, as (S2) states? There is ample evidence that they are not. The chemical atomic theory in the 18th century was so remarkably unsuccessful that most chemists abandoned it in favor of a more phenomenological, elective affinity chemistry. The Proutian theory that the atoms of heavy elements

are composed of hydrogen atoms had, through most of the 19th century, a strikingly unsuccessful career, confronted by a long string of apparent refutations. The Wegenerian theory that the continents are carried by large subterranean objects moving laterally across the earth's surface was, for some thirty years in the recent history of geology, a strikingly unsuccessful theory until, after major modifications, it became the geological orthodoxy of the 1960s and 1970s. Yet all of these theories postulated basic entities which (according to Putnam's 'principle of charity') genuinely exist.

The realist's claim that we should expect referring theories to be empirically successful is simply false. And, with a little reflection, we can see good reasons why it should be. To have a genuinely referring theory is to have a theory which "cuts the world at its joints", a theory which postulates entities of a kind that really exist. But a genuinely referring theory need not be such that all—or even most—of the specific claims it makes about the properties of those entities and their modes of interaction are true. Thus, Dalton's theory makes many claims about atoms which are false; Bohr's early theory of the electron was similarly flawed in important respects. Contra-(S2), genuinely referential theories need not be strikingly successful, since such theories may be 'massively false' (i.e., have far greater falsity content than truth content).

(S2) is so patently false that it is difficult to imagine that the realist need be committed to it. But what else will do? The (Putnamian) realist wants attributions of reference to a theory's terms to function in an explanation of that theory's success. The simplest and crudest way of doing that involves a claim like (S2). A less outrageous way of achieving the same end would involve the weaker,

S2' A theory whose terms refer will usually (but not always) be successful.

Isolated instances of referring but unsuccessful theories, sufficient to refute (S2), leave (S2') unscathed. But, if we were to find a broad range of referring but unsuccessful theories, that would be evidence against (S2'). Such theories can be generated at will. For instance, take any set of terms which one believes to be genuinely referring. In any language rich enough to contain negation, it will be possible to construct indefinitely many unsuccessful theories, all of whose substantive terms are genuinely referring. Now, it is always open to the realist to claim that such 'theories' are not really theories at all, but mere conjunctions of isolated statements—lacking that sort of conceptual integration we associate with 'real' theories. Sadly a parallel argument can be made for genuine theories. Consider, for instance, how many inadequate versions of the atomic theory there were in the 2000 years of atomic 'speculating', before a genuinely successful theory emerged. Consider how many unsuccessful versions there were of the wave theory of light before the 1820s, when a successful wave theory first

emerged. Kinetic theories of heat in the seventeenth and eighteenth century, developmental theories of embryology before the late nineteenth century sustain a similar story. (S2'), every bit as much as (S2), seems hard to reconcile with the historical record.

As Richard Burian has pointed out to me (in personal communication), a realist might attempt to dispense with both of those theses and simply rest content with (S3) alone. Unlike (S2) and (S2'), (S3) is not open to the objection that referring theories are often unsuccessful, for it makes no claim that referring theories are always or generally successful. But (S3) has difficulties of its own. In the first place, it seems hard to square with the fact that the central terms of many relatively successful theories (e.g., aether theories, phlogistic theories) are evidently non-referring. I shall discuss this tension in detail below. More crucial for our purposes here is that (S3) is *not strong enough* to permit the realist to utilize reference to explain success. Unless genuineness of reference entails that all or most referring theories will be successful, then the fact that a theory's terms refer scarcely provides a convincing explanation of that theory's success. If, as (S3) allows, many (or even most) referring theories can be unsuccessful, how can the fact that a successful theory's terms refer be taken to explain why it is successful? (S3) may or may not be true; but in either case it arguably gives the realist no explanatory access to scientific success.

A more plausible construal of Putnam's claim that reference plays a role in explaining the success of science involves a rather more indirect argument. It might be said (and Putnam does say this much) that we can explain why a theory is successful by assuming that the theory is true or approximately true. Since a theory can only be true or nearly true (in any sense of those terms open to the realist) if its terms genuinely refer, it might be argued that reference gets into the act willy-nilly when we explain a theory's success in terms of its truth(like) status. On this account, reference is piggy-backed on approximate truth. The viability of this indirect approach is treated at length in section 4 below so I shall not discuss it here except to observe that if the only contact point between reference and success is provided through the medium of approximate truth, then the link between reference and success is extremely tenuous.

What about (S3), the realist's claim that success creates a rational presumption of reference? We have already seen that (S3) provides no explanation of the success of science, but does it have independent merits? The question specifically is whether the success of a theory provides a warrant for concluding that its central terms refer. Insofar as this is—as certain realists suggest—an empirical question, it requires us to inquire whether past theories which have been successful are ones whose central terms genuinely referred (according to the realist's own account of reference).

A proper empirical test of this hypothesis would require extensive sift-

ing of the historical record of a kind that is not possible to perform here. What I can do is to mention a range of once successful, but (by present lights) non-referring, theories. A fuller list will come later (see section 5), but for now we shall focus on a whole family of related theories, namely, the subtle fluids and aethers of 18th and 19th century physics and chemistry.

Consider specifically the state of aetherial theories in the 1830s and 1840s. The electrical fluid, a substance which was generally assumed to accumulate on the surface rather than permeate the interstices of bodies, had been utilized to explain *inter alia* the attraction of oppositely charged bodies, the behavior of the Leyden jar, the similarities between atmospheric and static electricity and many phenomena of current electricity. Within chemistry and heat theory, the caloric aether had been widely utilized since Boerhaave (by, among others, Lavoisier, Laplace, Black, Rumford, Hutton, and Cavendish) to explain everything from the role of heat in chemical reactions to the conduction and radiation of heat and several standard problems of thermometry. Within the theory of light, the optical aether functioned centrally in explanations of reflection, refraction, interference, double refraction, diffraction and polarization. (Of more than passing interest, optical aether theories had also made some very startling predictions, e.g., Fresnel's prediction of a bright spot at the center of the shadow of a circular disc; a surprising prediction which, when tested, proved correct. If that does not count as empirical success, nothing does!) There were also gravitational (e.g., LeSage's) and physiological (e.g., Hartley's) aethers which enjoyed some measure of empirical success. It would be difficult to find a family of theories in this period which were as successful as aether theories; compared to them, 19th century atomism (for instance), a genuinely referring theory (on realist accounts), was a dismal failure. Indeed, on any account of empirical success which I can conceive of, non-referring 19th-century aether theories were more successful than contemporary, referring atomic theories. In this connection, it is worth recalling the remark of the great theoretical physicist, J. C. Maxwell, to the effect that the aether was better confirmed than any other theoretical entity in natural philosophy!

What we are confronted by in 19th-century aether theories, then, is a wide variety of once successful theories, whose central explanatory concept Putnam singles out as a prime example of a non-referring one (Putnam 1978, p. 22). What are (referential) realists to make of this historical case? On the face of it, this case poses two rather different kinds of challenges to realism: (1) it suggests that (S3) is a dubious piece of advice in that *there can be* (and have been) *highly successful theories some central terms of which are non-referring*; and (2) it suggests that *the realist's claim that he can explain why science is successful is false at least insofar as a part of the historical success of science has been success exhibited by theories whose central terms did not refer*.

But perhaps I am being less than fair when I suggest that the realist is committed to the claim that *all* the central terms in a successful theory refer. It is possible that when Putnam, for instance, says that "terms in a mature [or successful] science typically refer" (Putnam 1978, p. 20), he only means to suggest that *some* terms in a successful theory or science genuinely refer. Such a claim is fully consistent with the fact that certain other terms (e.g., 'aether') in certain successful, mature sciences (e.g., 19th-century physics) are nonetheless non-referring. Put differently, the realist might argue that the success of a theory warrants the claim that at least some (but not necessarily all) of its central concepts refer.

Unfortunately, such a weakening of (S3) entails a theory of evidential support which can scarcely give comfort to the realist. After all, part of what separates the realist from the positivist is the former's belief that the evidence for a theory is evidence for *everything* which the theory asserts. Where the stereotypical positivist argues that the evidence selectively confirms only the more 'observable' parts of a theory, the realist generally asserts (in the language of Boyd) that:

the sort of evidence which ordinarily counts in favor of the acceptance of a scientific law or theory is, ordinarily, evidence for the (at least approximate) truth of the law or theory as an account of the causal relations obtaining between the entities ["observational or theoretical"] quantified over in the law or theory in question. (Boyd 1973, p. 1)⁵

For realists such as Boyd, either all parts of a theory (both observational and non-observational) are confirmed by successful tests or none are. In general, realists have been able to utilize various holistic arguments to insist that it is not merely the lower-level claims of a well-tested theory which are confirmed but its deep-structural assumptions as well. This tactic has been used to good effect by realists in establishing that inductive support 'flows upward' so as to authenticate the most 'theoretical' parts of our theories. Certain latter-day realists (e.g., Glymour) want to break out of this holist web and argue that certain components of theories can be 'directly' tested. This approach runs the very grave risk of undercutting what the realist desires most: a rationale for taking our deepest-structure theories seriously, and a justification for linking reference and success. After all, if the tests to which we subject our theories only test *portions* of those theories, then even highly successful theories may well have central terms which are non-referring and central tenets which, because untested, we have no grounds for believing to be approximately true. Under those circumstances, a theory might be highly successful and yet contain important constituents which were patently false. Such a state of affairs would wreak havoc with the realist's presumption (R1) that success betokens approximate truth. In short, to be less than a holist about theory testing is

to put at risk precisely that predilection for deep-structure claims which motivates much of the realist enterprise.

There is, however, a rather more serious obstacle to this weakening of referential realism. It is true that by weakening (S3) to only certain terms in a theory, one would immunize it from certain obvious counter-examples. But such a maneuver has debilitating consequences for other central realist theses. Consider the realist's thesis (R3) about the retentive character of inter-theory relations (discussed below in detail). The realist both recommends as a matter of policy and claims as a matter of fact that successful theories are (and should be) rationally replaced only by theories which preserve reference for the central terms of their successful predecessors. The rationale for the normative version of this retentionist doctrine is that the terms in the earlier theory, *because it was successful, must* have been referential and thus a constraint on any successor to that theory is that reference should be retained for such terms. This makes sense just in case success provides a blanket warrant for presumption of reference. But if (S3) were weakened so as to say merely that it is reasonable to assume that *some* of the terms in a successful theory genuinely refer, then the realist would have no rationale for his retentive theses (variants of R3), which have been a central pillar of realism for several decades.⁶

Something apparently has to give. A version of (S3) strong enough to license (R3) seems incompatible with the fact that many successful theories contain non-referring central terms. But any weakening of (S3) dilutes the force of, and removes the rationale for, the realist's claims about convergence, retention and correspondence in inter-theory relations.⁷ If the realist once concedes that some unspecified set of the terms of a successful theory may well not refer, then his proposals for restricting "the class of candidate theories" to those which retain reference for the *prima facie* referring terms in earlier theories is without foundation. (Putnam 1975, p. 22)

More generally, we seem forced to say that such linkages as there are between reference and success are rather murkier than Putnam's and Boyd's discussions would lead us to believe. If the realist is going to make his case for CER, it seems that it will have to hinge on approximate truth, (R1), rather than reference, (R2).

4 | Approximate Truth and Success: the 'Downward Path'

Ignoring the referential turn among certain recent realists, most realists continue to argue that, at bottom, epistemic realism is committed to the view that successful scientific theories, even if strictly false, are nonetheless

'approximately true' or 'close to the truth' or 'verisimilar'.⁸ The claim generally amounts to this pair:

- T1 if a theory is approximately true, then it will be explanatorily successful; and
- T2 if a theory is explanatorily successful, then it is probably approximately true.

What the realist would *like* to be able to say, of course, is:

- T1' if a theory is true, then it will be successful.

(T1') is attractive because self-evident. But most realists balk at invoking (T1') because they are (rightly) reluctant to believe that we can reasonably presume of any given scientific theory that it is true. If all the realist could explain was the success of theories which were true *simpliciter* [unconditionally], his explanatory repertoire would be acutely limited. As an attractive move in the direction of broader explanatory scope, (T1) is rather more appealing. After all, presumably many theories which we believe to be false (e.g., Newtonian mechanics, thermodynamics, wave optics) were—and still are—highly successful across a broad range of applications.

Perhaps, the realist evidently conjectures, we can find an *epistemic* account of that pragmatic success by assuming such theories to be 'approximately true'. But we must be wary of this potential sleight of hand. It may be that there is a connection between success and approximate truth; *but if there is such a connection it must be independently argued for*. The acknowledgedly uncontroversial character of (T1') must not be surreptitiously invoked—as it sometimes seems to be—in order to establish (T1). When (T1')'s antecedent is appropriately weakened by speaking of approximate truth, it is by no means clear that (T1) is sound.

Virtually all the proponents of epistemic realism take it as unproblematic that if a theory were approximately true, it would deductively follow that the theory would be a relatively successful predictor and explainer of observable phenomena. Unfortunately, few of the writers of whom I am aware have defined what it means for a statement or theory to be 'approximately true'. Accordingly, it is impossible to say whether the alleged entailment is genuine. This reservation is more than perfunctory. Indeed, on the best known account of what it means for a theory to be approximately true, it does *not* follow that an approximately true theory will be explanatorily successful.

Suppose, for instance, that we were to say in a Popperian vein that a theory, T_1 , is approximately true if its truth content is greater than its falsity content, i.e.,

$$Ct_T(T_1) \gg Ct_F(T_1).^9$$

(Where $Ct_T(T_1)$ is the cardinality of the set of true sentences entailed by T_1 and $Ct_F(T_1)$ is the cardinality of the set of false sentences entailed by T_1 .) When approximate truth is so construed, it does *not* logically follow that an arbitrarily selected class of a theory's entailments (namely, some of its observable consequences) will be true. Indeed, it is entirely conceivable that a theory might be approximately true in the indicated sense and yet be such that *all* of its thus far tested consequences are *false*.¹⁰

Some realists concede their failure to articulate a coherent notion of approximate truth or verisimilitude, but insist that this failure in no way compromises the viability of (T1). Newton-Smith, for instance, grants that "no one has given a satisfactory analysis of the notion of verisimilitude" (Newton-Smith 1981, p. 197), but insists that the concept can be legitimately invoked "even if one cannot at the time give a philosophically satisfactory analysis of it." He quite rightly points out that many scientific concepts were explanatorily useful long before a philosophically coherent analysis was given for them. But the analogy is unseemly, for what is being challenged is not whether the concept of approximate truth is philosophically rigorous but rather whether it is even clear enough for us to ascertain whether it entails what it purportedly explains. Until someone provides a clearer analysis of approximate truth than is now available, it is not even clear whether truth-likeness would explain success, let alone whether, as Newton-Smith insists, "the concept of verisimilitude is *required* in order to give a satisfactory theoretical explanation of an aspect of the scientific enterprise." If the realist would de-mystify the 'miraculousness' (Putnam) or the 'mysteriousness' (Newton-Smith¹¹) of the success of science, he needs more than a promissory note that somehow, someday, someone will show that approximately true theories must be successful theories.¹²

Whether there is some definition of approximate truth which does indeed entail that approximately true theories will be predictively successful (and yet still probably false) is not clear.¹³ What can be said is that, promises to the contrary notwithstanding, *none* of the proponents of realism has yet articulated a coherent account of approximate truth which entails that approximately true theories will, across the range where we can test them, be successful predictors. Further difficulties abound. Even if the realist had a semantically adequate characterization of approximate or partial truth, and even if that semantics entailed that most of the consequences of an approximately true theory would be true, he would still be without any criterion that would *epistemically* warrant the ascription of approximate truth to a theory. As it is, the realist seems to be long on intuitions and short on either a semantics or an epistemology of approximate truth.

These should be urgent items on the realists' agenda since, until we

have a coherent account of what approximate truth is, central realist theses like (R1), (T1) and (T2) are just so much mumbo-jumbo.

5 | Approximate Truth and Success: the ‘Upward Path’

Despite the doubts voiced in section 4, let us grant for the sake of argument that if a theory is approximately true, then it will be successful. Even granting (T1), is there any plausibility to the suggestion of (T2) that explanatory success can be taken as a rational warrant for a judgment of approximate truth? The answer seems to be “no”.

To see why, we need to explore briefly one of the connections between ‘genuinely referring’ and being ‘approximately true’. However the latter is understood, I take it that *a realist would never want to say that a theory was approximately true if its central theoretical terms failed to refer*. If there were nothing like genes, then a genetic theory, no matter how well confirmed it was, would not be approximately true. If there were no entities similar to atoms, no atomic theory could be approximately true; if there were no sub-atomic particles, then no quantum theory of chemistry could be approximately true. In short, a necessary condition—especially for a scientific realist—for a theory being close to the truth is that its central explanatory terms genuinely refer. (An *instrumentalist*, of course, could countenance the weaker claim that a theory was approximately true so long as its directly testable consequences were close to the observable values. But as I argued above, the realist must take claims about approximate truth to refer alike to the observable and the deep-structural dimensions of a theory.)

Now, what the history of science offers us is a plethora of theories which were both successful and (so far as we can judge) non-referential with respect to many of their central explanatory concepts. I discussed earlier one specific family of theories which fits this description. Let me add a few more prominent examples to the list:

- the crystalline spheres of ancient and medieval astronomy;
- the humoral theory of medicine;
- the effluvial theory of static electricity;
- ‘catastrophist’ geology, with its commitment to a universal (Noachian) deluge;
- the phlogiston theory of chemistry;
- the caloric theory of heat;
- the vibratory theory of heat;
- the vital force theories of physiology;

- the electromagnetic aether;
- the optical aether;
- the theory of circular inertia;
- theories of spontaneous generation.

This list, which could be extended *ad nauseam*, involves in every case a theory which was once successful and well confirmed, but which contained central terms which (we now believe) were non-referring. Anyone who imagines that the theories which have been successful in the history of science have also been, with respect to their central concepts, genuinely referring theories has studied only the more 'whiggish' versions of the history of science (i.e., the ones which recount only those past theories which are referentially similar to currently prevailing ones).

It is true that proponents of CER sometimes hedge their bets by suggesting that their analysis applies exclusively to 'the mature sciences' (e.g., Putnam and Krajewski). This distinction between mature and immature sciences proves convenient to the realist since he can use it to dismiss any *prima facie* counter-example to the empirical claims of CER on the grounds that the example is drawn from an 'immature' science. But this insulating manoeuvre is unsatisfactory in two respects. In the first place, it runs the risk of making CER vacuous since these authors generally define a mature science as one in which correspondence or limiting case relations obtain invariably between any successive theories in the science once it has passed 'the threshold of maturity'. Krajewski grants the tautological character of this view when he notes that "the thesis that there is [correspondence] among successive theories becomes, indeed, analytical" (1977, p. 91). Nonetheless, he believes that there is a version of the maturity thesis which "may be and must be tested by the history of science". That version is that "every branch of science crosses at some period the threshold of maturity". But the testability of this hypothesis is dubious at best. There is no historical observation which could conceivably *refute* it since, even if we discovered that no sciences yet possessed 'corresponding' theories, it could be maintained that eventually every science will become corresponding. It is equally difficult to *confirm* it since, even if we found a science in which corresponding relations existed between the latest theory and its predecessor, we would have no way of knowing whether that relation will continue to apply to subsequent changes of theory in that science. In other words, the much-vaunted empirical testability of realism is seriously compromised by limiting it to the mature sciences.

But there is a second unsavory dimension to the restriction of CER to the 'mature' sciences. The realists' avowed aim, after all, is to explain why science is successful: that is the 'miracle' which they allege the non-realists leave unaccounted for. The fact of the matter is that parts of science, including many 'immature' sciences, have been successful for a very

long time; indeed, many of the theories I alluded to above were empirically successful by any criterion I can conceive of (including fertility, intuitively high confirmation, successful prediction, etc.). If the realist restricts himself to explaining only how the 'mature' sciences work (and recall that very few sciences indeed are yet 'mature' as the realist sees it), then he will have completely failed in his ambition to explain why science in general is successful. Moreover, several of the examples I have cited above come from the history of mathematical physics in the last century (e.g., the electromagnetic and optical aethers) and, as Putnam himself concedes, "*physics* surely counts as a 'mature' science if any science does" (1978, p. 21). Since realists would presumably insist that many of the central terms of the theories enumerated above do not genuinely refer, it follows that none of those theories could be approximately true (recalling that the former is a necessary condition for the latter). Accordingly, cases of this kind cast very grave doubts on the plausibility of (T2), i.e., the claim that nothing succeeds like approximate truth.

I daresay that for every highly successful theory in the past of science which we now believe to be a genuinely referring theory, one could find half a dozen once successful theories which we now regard as substantially non-referring. If the proponents of CER are the empiricists they profess to be about matters epistemological, cases of this kind and this frequency should give them pause about the well-foundedness of (T2).

But we need not limit our counter-examples to non-referring theories. There were many theories in the past which (so far as we can tell) were both genuinely referring and empirically successful which we are nonetheless loath to regard as approximately true. Consider, for instance, virtually all those geological theories prior to the 1960s which denied any lateral motion to the continents. Such theories were, by any standard, highly successful (and apparently referential); but would anyone today be prepared to say that their constituent theoretical claims—committed as they were to laterally stable continents—are almost true? Is it not the fact of the matter that structural geology was a successful science between (say) 1920 and 1960, even though geologists were fundamentally mistaken about many—perhaps even most—of the basic mechanisms of tectonic construction? Or what about the chemical theories of the 1920s which assumed that the atomic nucleus was structurally homogenous? Or those chemical and physical theories of the late 19th century which explicitly assumed that matter was neither created nor destroyed? I am aware of no sense of approximate truth (available to the realist) according to which such highly successful, but evidently false, theoretical assumptions could be regarded as 'truthlike'.

More generally, the realist needs a riposte to the *prima facie* plausible claim that there is no necessary connection between increasing the accuracy of our deep-structural characterizations of nature and improvements at the level of phenomenological explanations, predictions and

manipulations. It *seems* entirely conceivable intuitively that the theoretical mechanisms of a new theory, T_2 , might be closer to the mark than those of a rival T_1 and yet T_1 might be more accurate at the level of testable predictions. In the absence of an argument that greater correspondence at the level of unobservable claims is more likely than not to reveal itself in greater accuracy at the experimental level, one is obliged to say that the realist's hunch that increasing deep-structural fidelity must manifest itself pragmatically in the form of heightened experimental accuracy has yet to be made cogent. (Equally problematic, of course, is the inverse argument to the effect that increasing experimental accuracy betokens greater truthlikeness at the level of theoretical, i.e., deep-structural, commitments.)^{*} . . .

7 | The Realists' Ultimate 'Petitio Principii'

It is time to step back a moment from the details of the realists' argument to look at its general strategy. Fundamentally, the realist is utilizing, as we have seen, an abductive inference which proceeds from the success of science to the conclusion that science is approximately true, verisimilar, or referential (or any combination of these). This argument is meant to show the sceptic that theories are not ill-gotten, the positivist that theories are not reducible to their observational consequences, and the pragmatist that classical epistemic categories (e.g., 'truth', 'falsehood') are a relevant part of meta-scientific discourse.

It is little short of remarkable that realists would imagine that their critics would find the argument compelling. As I have shown elsewhere (1978), ever since antiquity critics of epistemic realism have based their scepticism upon a deep-rooted conviction that the fallacy of affirming the consequent is indeed fallacious. When Sextus or Bellarmine or Hume doubted that certain theories which saved the phenomena were warrantable as true, their doubts were based on a belief that the exhibition that a theory had some true consequences left entirely open the truth-status of the theory. Indeed, many non-realists have been non-realists precisely because they believed that false theories, as well as true ones, could have true consequences.

Now enters the new breed of realist (e.g., Putnam, Boyd and Newton-Smith) who wants to argue that epistemic realism can reasonably be pre-

^{*} The section that follows in Laudan's original paper—"6. Confusions about Convergence and Retention"—has been omitted. In it, Laudan criticizes abductive arguments for realism of sort II, namely those that conclude that (probably) earlier theories in a mature science are approximately true and their terms refer because later, more successful theories in the science preserve those earlier theories as limiting cases.

sumed to be true by virtue of the fact that it has true consequences. But this is a monumental case of begging the question. The non-realist refuses to admit that a *scientific* theory can be warrantably judged to be true simply because it has some true consequences. Such non-realists are not likely to be impressed by the claim that a *philosophical* theory like realism can be warranted as true because it arguably has some true consequences. If non-realists are chary about first-order abductions to avowedly true conclusions, they are not likely to be impressed by second-order abductions, particularly when, as I have tried to show above, the premises and conclusions are so indeterminate.

But, it might be argued, the realist is not out to convert the intransigent sceptic or the determined instrumentalist.¹⁴ He is perhaps seeking, rather, to show that realism can be tested like any other scientific hypothesis, and that realism is at least as well confirmed as some of our best scientific theories. Such an analysis, however plausible initially, will not stand up to scrutiny. I am aware of no realist who is willing to say that a *scientific* theory can be reasonably presumed to be true or even regarded as well confirmed just on the strength of the fact that its thus far tested consequences are true. Realists have long been in the forefront of those opposed to *ad hoc* and *post hoc* theories. Before a realist accepts a scientific hypothesis, he generally wants to know whether it has explained or predicted more than it was devised to explain; he wants to know whether it has been subjected to a battery of controlled tests; whether it has successfully made novel predictions; whether there is independent evidence for it.

What, then, of realism itself as a 'scientific' hypothesis?¹⁵ Even if we grant (contrary to what I argued in section 4) that realism entails and thus explains the success of science, ought that (hypothetical) success warrant, by the realist's own construal of scientific acceptability, the acceptance of realism? Since realism was devised in order to explain the success of science, it remains purely *ad hoc* with respect to that success. If realism has made some novel predictions or been subjected to carefully controlled tests, one does not learn about it from the literature of contemporary realism. At the risk of apparent inconsistency, the realist repudiates the instrumentalist's view that saving the phenomena is a significant form of evidential support while endorsing realism itself on the transparently instrumentalist grounds that it is confirmed by those very facts it was invented to explain. No proponent of realism has sought to show that realism satisfies those stringent empirical demands which the realist himself minimally insists on when appraising scientific theories. The latter-day realist often calls realism a 'scientific' or 'well-tested' hypothesis, but seems curiously reluctant to subject it to those controls which he otherwise takes to be a *sine qua non* for empirical well-foundedness.

8 | Conclusion

The arguments and cases discussed above seem to warrant the following conclusions:

- 1 The fact that a theory's central terms refer does not entail that it will be successful; and a theory's success is no warrant for the claim that all or most of its central terms refer.
- 2 The notion of approximate truth is presently too vague to permit one to judge whether a theory consisting entirely of approximately true laws would be empirically successful; what is clear is that a theory may be empirically successful even if it is not approximately true.
- 3 Realists have no explanation whatever for the fact that many theories which are not approximately true and whose 'theoretical' terms seemingly do not refer are nonetheless often successful.
- 4 The convergentist's assertion that scientists in a 'mature' discipline usually preserve, or seek to preserve, the laws and mechanisms of earlier theories in later ones is probably false; his assertion that when such laws are preserved in a successful successor, we can explain the success of the latter by virtue of the truthlikeness of the preserved laws and mechanisms, suffers from all the defects noted above confronting approximate truth.
- 5 Even if it could be shown that referring theories and approximately true theories would be successful, the realists' argument that successful theories are approximately true and genuinely referential takes for granted precisely what the non-realist denies (namely, that explanatory success betokens truth).
- 6 It is not clear that acceptable theories either *do* or *should* explain why their predecessors succeeded or failed. If a theory is better supported than its rivals and predecessors, then it is not epistemically decisive whether it explains why its rivals worked.
- 7 If a theory has once been falsified, it is unreasonable to expect that a successor should retain either all of its content or its confirmed consequences or its theoretical mechanisms.
- 8 Nowhere has the realist established—except by fiat—that non-realist epistemologists lack the resources to explain the success of science.

With these specific conclusions in mind, we can proceed to a more global one: it is not yet established—Putnam, Newton-Smith and Boyd notwithstanding—that realism can explain *any* part of the success of science. What is very clear is that realism *cannot*, even by its own lights, explain the success of those many theories whose central terms have evi-

dently not referred and whose theoretical laws and mechanisms were not approximately true. The inescapable conclusion is that insofar as many realists are concerned with explaining how science works and with assessing the adequacy of their epistemology by that standard, they have thus far failed to explain very much. Their epistemology is confronted by anomalies which seem beyond its resources to grapple with.

It is important to guard against a possible misinterpretation of this essay. *Nothing* I have said here refutes the possibility in principle of a realistic epistemology of science. To conclude as much would be to fall prey to the same inferential prematurity with which many realists have rejected in principle the possibility of explaining science in a non-realist way. My task here is, rather, that of reminding ourselves that there is a difference between wanting to believe something and having good reasons for believing it. All of us would like realism to be true; we would like to think that science works because it has got a grip on how things really are. But such claims have yet to be made out. Given the *present* state of the art, it can only be wish fulfilment that gives rise to the claim that realism, and realism alone, explains why science works.¹⁶

■ | Notes

1. Putnam, evidently following Boyd, sums up (R1) to (R3) in these words:

"1) Terms in a mature science typically *refer*.

2) The laws of a theory belonging to a mature science are typically approximately true . . . I will only consider [new] theories . . . which have this property—[they] contain the [theoretical] laws of [their predecessors] as a limiting case" (1978, pp. 20–21).

2. Putnam insists, for instance, that if the realist is wrong about theories being referential, then "the success of science [is] a miracle". (Putnam 1975, p. 73).

3. Boyd remarks: "scientific realism offers an *explanation* for the legitimacy of ontological commitment to theoretical entities" (Putnam 1978, note 2, p. 20). It allegedly does so by explaining why theories containing theoretical entities work so well: because such entities genuinely exist. [The quotation is from Boyd's (unpublished) "Realism and Scientific Epistemology."]

4. Whether one utilizes Putnam's earlier or later versions of realism is irrelevant for the central arguments of this essay.

5. See also p. 3: "experimental evidence for a theory is evidence for the truth of even its non-observational laws". See also (Sellars 1963, p. 97).

6. A caveat is in order here. *Even if* all the central terms in some theory refer, it is not obvious that every rational successor to that theory must preserve all the referring terms of its predecessor. One can easily imagine circumstances when the new theory is preferable to the old one even though the range of application of the new theory is less broad than the old. When the range is so restricted, it may

well be entirely appropriate to drop reference to some of the entities which figured in the earlier theory.

7. For Putnam and Boyd both "it will be a *constraint* on T_2 [i.e., any new theory in a domain] . . . that T_2 must have this property, the property that *from its standpoint* one can assign referents to the terms of T_1 [i.e., an earlier theory in the same domain]." (Putnam 1978, p. 22). For Boyd, see (1973, p. 8): "new theories should, *prima facie*, resemble current theories with respect to their accounts of causal relations among theoretical entities".

8. For just a small sampling of this view, consider the following: "The claim of a realist ontology of science is that the only way of explaining why the models of science function so successfully . . . is that they approximate in some way the structure of the object" (McMullin 1970, pp. 63–64); "the continued success [of confirmed theories] can be *explained* by the hypothesis that they are in fact close to the truth . . ." (Niiniluoto 1980, p. 448); the claim that "the laws of a theory belonging to a mature science are typically approximately *true* . . . [provides] an *explanation* of the behavior of scientists and the success of science" (Putnam 1978, pp. 20–21). Smart, Sellars, and Newton-Smith, among others, share a similar view.

9. Although Popper is generally careful not to assert that actual historical theories exhibit ever increasing truth content (for an exception, see his (1963, p. 220)), other writers have been more bold. Thus, Newton-Smith writes that "the historically generated sequence of theories of a mature science is a sequence in which succeeding theories are increasing in truth content without increasing in falsity content." [See Newton-Smith 1981, p. 184. Laudan is quoting from a draft of Newton-Smith's book, the wording of which differs from the published version.]

10. On the more technical side, Niiniluoto has shown that a theory's degree of corroboration co-varies with its "estimated verisimilitude" (1977, pp. 121–147 and 1980). Roughly speaking, 'estimated truthlikeness' is a measure of how closely (the content of) a theory corresponds to *what we take to be* the best conceptual systems that we so far have been able to find (1980, pp. 443ff.). If Niiniluoto's measures work, it follows from the above-mentioned co-variance that an empirically successful theory will have a high degree of estimated truthlikeness. But because estimated truthlikeness and genuine verisimilitude are not necessarily related (the former being parasitic on existing evidence and available conceptual systems), it is an open question whether—as Niiniluoto asserts—the continued success of highly confirmed theories can be *explained* by the hypothesis that they in fact are close to the truth at least in the relevant respects. Unless I am mistaken, this remark of his betrays a confusion between 'true verisimilitude' (to which we have no epistemic access) and 'estimated verisimilitude' (which is accessible but non-epistemic).

11. Newton-Smith claims that the increasing predictive success of science through time "would be totally mystifying . . . if it were not for the fact that theories are capturing more and more truth about the world" (Newton-Smith 1981, p. 196).

12. I must stress again that I am *not* denying that there *may* be a connection between approximate truth and predictive success. I am only observing that until the realists show us what that connection is, they should be more reticent than they are about claiming that realism can explain the success of science.

13. A *non-realist* might argue that a theory is approximately true just in case all its *observable* consequences are true or within a specified interval from the true value. Theories that were "approximately true" in this sense would indeed be demonstrably successful. But the realist's (otherwise commendable) commitment to taking seriously the theoretical claims of a theory precludes him from utilizing any such construal of approximate truth, since he wants to say that the theoretical as well as the observational consequences are approximately true.

14. I owe the suggestion of this realist response to Andrew Lugg.

15. I find Putnam's views on the 'empirical' or 'scientific' character of realism rather perplexing. At some points, he seems to suggest that realism is both empirical and scientific. Thus, he writes: "If realism is an explanation of this fact [namely, that science is successful], realism must itself be an over-arching scientific hypothesis" (1978, p. 19). Since Putnam clearly maintains the antecedent, he seems committed to the consequent. Elsewhere he refers to certain realist tenets as being "our highest level empirical generalizations about knowledge" (p. 37). He says moreover that realism "could be false", and that "facts are relevant to its support (or to criticizing it)" (pp. 78–79). Nonetheless, for reasons he has not made clear, Putnam wants to deny that realism is either scientific or a hypothesis (p. 79). How realism can consist of doctrines which 1) explain facts about the world, 2) are empirical generalizations about knowledge, and 3) can be confirmed or falsified by evidence and yet be neither scientific nor hypothetical is left opaque.

16. I am indebted to all of the following for clarifying my ideas on these issues and for saving me from some serious errors: Peter Achinstein, Richard Burian, Clark Glymour, Adolf Grünbaum, Gary Gutting, Allen Janis, Lorenz Krüger, James Lennox, Andrew Lugg, Peter Machamer, Nancy Maull, Ernan McMullin, Ilkka Niiniluoto, Nicholas Rescher, Ken Schaffner, John Worrall, Stephen Wykstra.

■ | References

- Boyd, R. (1973), "Realism, Underdetermination, and a Causal Theory of Evidence", *Noûs* 7: 1–12.
- Krajewski, W. (1977), *Correspondence Principle and Growth of Science*. Dordrecht: Reidel.
- Laudan, L. (1978), "Ex-Huming Hacking", *Erkenntnis* 13: 417–35.
- McMullin, Ernan (1970), "The History and Philosophy of Science: A Taxonomy", Stuewer, R. (ed.), *Minnesota Studies in the Philosophy of Science* V: 12–67. Minneapolis: University of Minnesota Press.
- Newton-Smith, W. (1981), *The Rationality of Science*. London: Routledge & Kegan Paul.

Niiniluoto, Ilkka (1977), "On the Truthlikeness of Generalizations", Butts, R. and Hintikka, J. (eds.), *Basic Problems in Methodology and Linguistics*: 121–147. Dordrecht: Reidel.

Niiniluoto, Ilkka (1980), "Scientific Progress", *Synthese* 45: 427–62.

Popper, K. (1963), *Conjectures and Refutations*. London: Routledge & Kegan Paul.

Putnam, H. (1975), *Mathematics, Matter and Method*, Vol. 1. Cambridge: Cambridge University Press.

Putnam, H. (1978), *Meaning and the Moral Sciences*. London: Routledge & Kegan Paul.

Sellars, W. (1963), *Science, Perception and Reality*. New York: The Humanities Press.

Explaining the Success of Science

Karl Popper has steadfastly held that the success of science is not to be explained; it is a miracle. '... [N]o theory of knowledge', he says, 'should attempt to explain why we are successful in our attempts to explain things.'¹ And even though '... science has been miraculously successful ...', as he puts it, '[t]his strange fact cannot be explained.'² Consistency with his other views requires him, no doubt, to disavow any presupposition that a scientific theory is likely to be true. Yet explanations of the success of science often make that very assumption; a theory's *success* is explained by assuming that the theory is *true*. Hence Popper's quandary. But throwing up our hands in despair or embracing miracles seem neither the heroic nor the reasonable thing to do. I have nothing heroic to offer by way of accounting for the success of science either, but I will try a moderately reasonable stab at it. My proposal, however, will not be all that different from Popper's point of view.

Before proceeding further, something should be said about the term 'success'. There are several ways in which science is an overachiever: Its technological accomplishments are undeniable: it's very handy for building bridges and curing diseases. It is a glorious entertainer: many of us would rather curl up in bed with a good piece of popular physics than with any novel. And science has also been a great success at extracting tax dollars from us all. (I do not say that cynically; I would gladly pay more.)

By calling science successful I do not mean that everything that is called science is successful, only that many presently accepted theories are. And by calling these theories successful I simply mean that

1 They are able to organize and unify a great variety of known phenomena.

- 2 This ability to systematize the empirical data is more extensive now than it was for previous theories.
- 3 A statistically significant number of novel predictions pan out; that is, our theories get more predictions right than mere guessing would allow.

This, I think, is roughly what is involved in the normal use of the phrase 'the success of science', and I am simply following tradition here. At any rate these are the senses of success that I shall be dealing with. Even though they are common ingredients, they are not, however, always clearly distinguished by writers on this topic.

Before getting on to the main arguments I should also make the real motivation clear. Few concerned with this question care why science is successful *per se*. What they really care about are the ontological consequences of the various explanations. Realists, for example, think that if they can explain the success of a theory by appeal to the *truth* (or approximate truth) of that theory, then the ontological issue will be decided in their favour. Anti-realists, on the other hand, propose rival accounts which they see as ontologically innocuous. Since what is really at stake is the ontological question, it is best if I set the stage in terms of realists vs anti-realists.

■ | Miracles, Darwin, and 'The Truth'

The thing to be explained is the success of science and the way realists often explain this fact is by claiming that theories are true, or at least approximately true, and that any conclusion deduced from true premisses must itself be true. So the assumption that theories are (approximately) true explains the success of those theories. Realism, as Hilary Putnam puts it, is the only explanation which doesn't make the success of science a miracle. J. J. C. Smart states the case this way:

If the phenomenalist about theoretical entities is correct, we must believe in a *cosmic coincidence*. That is, if this is so, statements about electrons, etc., are of only instrumental value: they simply enable us to predict phenomena on the level of galvanometers and cloud chambers. They do nothing to remove the *surprising character* of these phenomena. . . . Is it not odd that the phenomena of the world should be such as to make a purely instrumental theory true? On the other hand, if we interpret a theory in a realist way, then we have no need for such a cosmic coincidence: it is not surprising that galvanometers and cloud chambers behave in the sort of way they do, for if there really are electrons, etc., this is just what we should expect. A lot of surprising facts no longer seem surprising.³

We can reconstruct the argument which lies buried in this passage in the way which makes it seem quite reasonable and convincing.

- 1 Conclusion O can be deduced from theory T.
- 2 O is observed to be the case.
- 3 If T is true then the argument for O is *sound* and so O *had* to be true.
- 4 If T is false then the argument for O is *merely valid* and the probability of the arbitrary consequence O being true is very small. (I.e., it would be a miracle if O were true.)

-
- ∴ The argument for O is probably sound.
 ∴ T is probably true. (That is, even T's theoretical statements are probably true.)

This argument uses the realist's explanation of the success of science to draw ontological morals. Let us contrast it with the rival Darwinian view of some anti-realists. The latter goes something like this: Just as there are a great many species struggling for existence, so too, a great many theories have been proposed. And just as species which are not adapted to their environment become extinct, so too, theories which are not making true observational predictions are dropped. The belief that our theories might be true, or even approximately true, is an illusion. It is similar to the illusion that Darwin undermined, that species are evolving *toward some goal*. Van Fraassen, to pick the most prominent recent anti-realist, gives a Darwinian account of the success of science in *The Scientific Image*:

I can best make the point by contrasting two accounts of the mouse who runs from its enemy, the cat. St. Augustine . . . provided an intentional explanation: the mouse *perceives that* the cat is its enemy, hence the mouse runs. What is postulated here is the 'adequacy' of the mouse's thought to the order of nature: the relation of enmity is correctly reflected in his mind. But the Darwinist says: Do not ask *why* the *mouse* runs from its enemy. Species which did not cope with their natural enemies no longer exist. That is why there are only ones who do.

And so, continues van Fraassen:

In just the same way, I claim that the success of current scientific theories is no miracle. It is not even surprising to the scientific (Darwinist) mind. For any scientific theory is born into a life of fierce competition, a jungle red in tooth and claw. Only the successful theories survive—the ones which *in fact* latched on to actual regularities in nature.⁴

'Truth' plays no role at all in the success of science for the Darwinian anti-realist. Yet for the realist it is the central explanatory factor. So here

we have the two main contenders, but could either of these explanations of the success of science be right?

■ | The Darwinian Answer

I characterized the success of science as having three ingredients. Van Fraassen's Darwinian explanation seems to account for two of these features, but not the third. He has an apparently adequate answer to the two questions, Why do theories get so much right?, and Why do newer theories get more right than the ones we have tossed out? The simple answer is that we have tossed out any theory which didn't organize, unify, and generally get a lot right; and we have tossed out theories which have done less well, comparatively, than others.

The third question seems to be still unanswered. Why do our theories make correct predictions more often than one could expect on the basis of mere chance? Here the Darwinian analogy breaks down since most species could not survive a radical change of environment, the analogue of the novel prediction.

There is also a more general problem with this Darwinian approach. It is a problem which stems from the empiricism of the anti-realists. An implicit assumption is that rational choice and success go hand in hand. On this assumption it is not surprising that science is successful in senses (1) and (2) since we choose theories, says the empiricist, on that very basis. This, I think, is not so. Success, as characterised by the anti-realist, is a totally empirical notion. But in reality theories are rationally evaluated on the basis of several other considerations besides empirical factors. I don't wish to argue here for any in particular, but let us suppose that conceptual, metaphysical, and aesthetic concerns play a role in actual theory choice. Consequently, it is not a trivial analytic truth that the rational thing to believe is also the most successful. Anyone who is not an extreme empiricist must concede that it is quite *possible* that the most rationally acceptable theory is not the most successful theory.

So even the Darwinian answers to (1) and (2) which above I tentatively conceded to be adequate are, in fact, not adequate after all. And (3), of course, remains entirely unexplained. The Darwinian account, linked to an empiricist methodology, yields a plausible account of two of the three aspects of success, but unlinked from this untenable methodology, it accounts for nothing.

■ | Realism and Reference

A belief common to scientific realists is that the succession of theories is getting closer to the truth. This belief may well be true (I hope it is), but it is often tied to a doctrine that says that the central terms of one theory refer to the same things as the central terms of its successor and predecessor theories. Moreover, the intuitive idea of getting-closer-to-the-truth will itself need fleshing out in the form of an explicit doctrine of verisimilitude. Unfortunately, there are terrible problems with both of these. Beliefs about the constancy of reference run afoul of the history of science, and the concept of verisimilitude is plagued with technical problems. Even a cursory glance at the past suggests that there is no royal road to the truth such as that implied by the convergence picture, and every explication of verisimilitude so far proposed has been a crashing failure.

Let's look at things now in some detail. Putnam gives a forthright version of the realist's explanation of the success of science in the following passage:

The positive argument for realism is that it is the only philosophy that doesn't make the success of science a miracle. That terms in mature theories typically refer (this formulation is due to Richard Boyd), that the theories accepted in a mature science are typically approximately true, that the same term can refer to the same thing even when it occurs in different theories—these statements are viewed by the scientific realist not as necessary truths but as part of the only scientific explanation of the success of science, and hence as part of any adequate scientific description of science and its relations to its objects.⁵

In the next section I will examine the idea that mature theories are 'typically approximately true' by looking at Newton-Smith's views since they are much more developed than Putnam's. This section will be devoted solely to examining the claim that 'terms in mature theories typically refer'. Let us begin by looking at a very simple theory:

T₁: JRB went to Dubrovnik in April.

For the sake of the argument, let us suppose that it is quite a successful theory (there were reports of his being seen there, etc.) and that all the terms in it refer. Is the fact that all the terms refer *sufficient* to explain why the theory is successful? The simplest consideration completely undermines this supposition. The following theory is very unsuccessful:

T₂: JRB did *not* go to Dubrovnik in April.

Yet all the terms just as surely refer.

Not all counter-examples are so artificial; historical illustrations of the problem abound. Consider the succession of atomistic theories; some were successful, but many were not. So clearly, having the term 'atom' in the theory does not lead to success even though (we believe) the term 'atom' refers.

Reference is not sufficient for success, but is it *necessary*? This, too, seems most unlikely. Phlogiston theories, caloric theories, aether theories, and numerous others have all had a definite heyday, yet by our present best guesses the central terms of these theories do not refer.

In the Putnam-Boyd explanation of the success of science there is a caveat. The term 'typically' is used; 'terms typically refer' and theories are 'typically approximately true'. This seems to leave one free to dismiss the occasional example such as phlogiston or caloric as a tolerable aberration. It would then appear to be a question of degree and consequently the historical case for or against this sort of realism is going to be rather difficult to establish.

One could seriously doubt that the historical cases will come out the way they expect, that is, with successful theories *typically* having terms which refer. But even if this should be the case with almost every theory, there still remains one great problem. A *single* example of a successful theory with at least one term which does not refer must count as a miracle. Thus, the success of the caloric theory of heat, by the lights of Putnam and Boyd, must rank with the raising of Lazarus from the dead. And what Priestly achieved with his phlogiston theory was no less an amazing feat than if he had turned water into wine. By weakening the claim to just saying that reference is *typical*, easy counter-examples drawn from the history of science might be avoided. But the cost is impossibly high: every untypical example is a miracle.

■ | Realism and Verisimilitude

It is time now to look at the other key idea in the Putnam-Boyd explanation of the success of science, the idea that theories are 'typically approximately true'. Unfortunately, neither Putnam nor Boyd have bothered to unpack this notion, so I will examine the similar but rather more developed views of William Newton-Smith instead.

Newton-Smith's approach to verisimilitude is a 'transcendental' one as he puts it. He too is looking for an explanation of what he sees is an undeniable fact: *science has made progress*. And how has this remarkable achievement come about? His realist answer is disarmingly simple: If our theories were getting closer to the truth then this is exactly what we should expect.⁶

To maintain a doctrine of increasing *verisimilitude*, or truth-likeness, is to maintain that the succession of past theories, up to the present, have been getting *closer to the truth*. There may be several respects in which later theories are better than earlier ones; they may be better predictors, more elegant, technologically more fruitful. But the one respect the realist cares about most is veracity; later theories, it is hoped and claimed, are better in this regard. Verisimilitude is an intuitive notion that most people subscribe to; but it is extremely problematic. The most famous instance of trying to come to grips with it, namely Popper's account, is a clear-cut failure. And unless someone is able to successfully explicate the notion soon, it is likely to have the same fate as such other intuitive notions as 'neutral observation' and 'simplicity'. It will be tossed on the junk pile of history.

There is one virtue of Newton-Smith's account of verisimilitude which needs stressing. Constancy of reference across successive theories is not required. The kind of problems phlogiston, caloric, and the aether present for the convergence account of Putnam and Boyd have no bearing on Newton-Smith's version. This is what makes his recent account interesting, initially promising, and worthy of special attention.

Let me now focus on some of the details of his account of verisimilitude. What is required, as he sees it,⁷ is an analysis of the notion which will then justify the crucial premiss in his argument. That is, he must show that, on unpacking, the concept of verisimilitude yields this: *An increase in verisimilitude implies the likeliness of an increase in observational success*. And he is quite right to worry about this, for in spite of its intuitive nature, we cannot count on the properties of *truth* carrying over for *truth-likeness*. The consequences of a true theory are true; but the consequences of a theory which is approximately true may not themselves be approximately true.

Before getting to his analysis of verisimilitude, we need to set the stage with Newton-Smith's characterization of a few key notions. A *theory* is the deductive closure of the postulates and appropriate auxiliary hypotheses; an *observational consequence* is a conditional, $p \rightarrow q$, where p is a statement of the observable initial conditions and q the observable final conditions; the consequences of a theory must be *recursively enumerable*, (i.e., can be mechanically produced in a sequence. Newton-Smith does not defend this dubious condition.) A theory *decides* p if it implies either p or its negation. The *content* of a theory is a fairly technical notion, but we can say roughly that a theory has more content than another if it decides more sentences. Since typically both will decide infinitely many sentences some technical complications in the definition are required. Imagine two theories, T_1 and T_2 , with their consequences recursively enumerated. The n th member of the sequence generated from T_1 either will or will not be decided by T_2 . We are to determine which it is. (Newton-Smith fails to note that given Church's theorem, this is not going to be generally pos-

sible.) This process is generalized and finally we are able to form the appropriate ratio from the sentences decided by the two theories. In this way Newton-Smith is able to define which theory has the greater content, and he is able to do so in a manner which seems to capture our intuitive requirements. Of course, the definition is based on an infinite sequence, but for practical purposes, greater content could be determined after a large, but finite, number of sentences have been examined.

The last important notion is that of *relative truth*. Consider again the theories T_1 and T_2 with their consequences enumerated recursively. After n terms there will be a number of truths and a number of falsehoods for each. The ratio of these numbers is the *truth ratio*. We then pick a third theory T_3 to appraise the truth values of the sentences in the sequence generated by T_1 and T_2 . (T_3 could be either from a God's eye point of view or it could be our presently held theory.) Newton-Smith then defines T_2 as having *greater truth relative to T_3* than T_1 has, if and only if the infinite sequence of ratios, which give the ratio of truths in T_1 to the truths in T_2 as judged by reference to T_3 , has a limit greater than $1/2$.^{*} Now we come to the main idea:

T_2 has greater *verisimilitude* than T_1 if and only if both: (1) the relative content of T_2 is equal to or greater than T_1 ; (2) T_2 has greater truth relative to T_3 than T_1 .⁸

So the rough idea is this: To have more verisimilitude is to say more about the world and to say more true things in doing so. Does this solve the initial problem which was to show the greater verisimilitude implied the likelihood of greater observational success? The answer, says Newton-Smith, is yes. Here is his argument: Pick an arbitrary sentence from T_2 which we will assume has greater verisimilitude than T_1 according to the definition. The chances of it being true, since it came from T_2 , are greater than the chances of some arbitrary sentence which comes from T_1 being true. And since the set of arbitrary sentences of T_2 include the observational sentences it follows that T_2 will likely have more observational successes.

This account of the notion of truth-likeness certainly has its attractions. It is not obviously plagued with the same problems which beset Popper's account; it is simple and elegant; and it satisfies several of our most basic intuitions about the concept. However, it still seems to be not entirely satisfactory, as a number of considerations show.

Is Newton-Smith's explanation good at accounting for all three senses of success? Not entirely. It is very good at accounting for (2) and (3). But it doesn't say why present theories get much right; it is perfectly compatible

There is an error in this definition, which Newton-Smith has since corrected. See the section "Verisimilitude and Success" in the commentary on this chapter.

with Newton-Smith's theory that our present beliefs organize the data poorly and generally get very little right. His theory guarantees that our present scientific theories do better than our past theories, and he can also explain why present theories get more novel predictions right than mere chance would allow for. But there is still one important sense of success left unexplained.

Second, consider the following situation:

The Truth: John is 180 cm tall.

Theory₁: John is 178 cm tall.

Theory₂: John is 179 cm tall.

These two theories say only one thing each. Intuitively, I think we should all agree that T_2 is closer to the truth. But on Newton-Smith's account we cannot say this. The two theories are both totally and equally false. The reason for this is simply that Newton-Smith takes the appropriate measure to be a comparison of the true sentences with the false ones. Individual sentences are taken to be one or the other, and never something in between. Yet a proper theory of verisimilitude would, I think, take this into consideration. And it would have to in order to do full justice to the history of science. Consider a succession of mini-theories which are really about nothing more than the value of some physical constant, say, Planck's constant, the gravitational constant, or the magnetic moment of the electron. Progress in such a case would be a sequence of claims which are becoming ever more accurate, though none as yet exactly right. Every claim as to what the value is would be false, yet surely it is correct to see this sequence as getting closer to the truth, if anything does.

A third problem that I see with Newton-Smith may be yet more serious. Historical considerations make his requirement of increasing content in the definition of greater verisimilitude implausible. Any event in the history of science where the domain shrunk, and there are several of them, will stand as a counterexample. Newton-Smith's requirement is that the later theory must have equal or greater content than the former. But this did not happen in the following example which most of us would likely consider a progressive move: Once there were theories which combined astronomy and astrology together, then a transition was made to purely astronomical theories. The earlier theories which combined both astronomical and astrological claims obviously said more about the world, so the later astronomical theories had less content. However we characterize truth-likeness, it must be compatible with such domain-shrinking transitions in the history of science. Newton-Smith's account is not.

■ | Is Hypothetico-Deductivism the Problem?

What about the style of Newton-Smith's argument which links greater verisimilitude with the likelihood of greater observational success? Anti-realists often decry the hypothetico-deductive (H-D) form of inference. That is, they reject arguments which go:

Theory \rightarrow Observation
 Observation

 \therefore (Probably) Theory

Given that they find H-D arguments unconvincing (claiming that it is a simple fallacy of affirming the consequent), why should anti-realists be persuaded to become realists by an argument that goes: Verisimilitude would explain greater observational success and there has been greater observational success; thus, there must be greater verisimilitude? The style is the same in both cases:

Greater verisimilitude \rightarrow Greater observational success
 Greater observational success

 \therefore (Probably) Greater verisimilitude

The anti-realist will simply say that the question has been begged. Some of us may like Newton-Smith's argument for verisimilitude and the realist approach in general, but then we *already* liked H-D inference. Laudan gives voice to this anti-realist sentiment when he writes:

... ever since antiquity critics of epistemic realism have based their scepticism upon a deep-rooted conviction that the fallacy of affirming the consequent is indeed fallacious. ... Now enters the new breed of realist ... who wants to argue that epistemic realism can reasonably be presumed to be true by virtue of the fact that it has true consequences. But this is a monumental case of begging the question.⁹

Can the blame for the failures to explain the success of science be pinned on H-D inference? At first glance the fight between realist and anti-realist over the success of science seems but a dressed up version of the old problem of induction. If there is no hope of solving that problem, then how can we hope to explain the success of science? The answer, I think, is that they are not really the same problem. If H-D reasoning were really the issue here it would be a problem for anti-realists too. But van

Fraassen, a paradigm anti-realist, relies on H-D inference regularly, as he must for instance, in the following type of argument:

T is empirically adequate	→	Observation O
Observation O		
<hr/>		
∴ (Probably) T is empirically adequate.		

Van Fraassen wants to go as little beyond the observable evidence as he can, but he does take some risks. He resists inferences to the truth but in accepting a theory as empirically adequate he recognizes the need for ampliative inference.

Similarly, Laudan, when he has his historian's hat on, says the shift to the H-D style of inference with Hartley and LeSage was a great step forward in the history of methodology.¹⁰ Before their work, the Newtonian tradition of doing science was based on the famous dictum, *hypotheses non fingo*; theories were to be deduced from the phenomena. The introduction of H-D [reasoning] in the 18th century marked a definite advance.

Anti-realists such as van Fraassen and Laudan are not *sceptics* about induction. They need and use inductive inference as much as realists do. If realists are committing a fallacy at the meta-level of explaining science, then so is everyone else (except perhaps Popper) at the theory level of explaining the world. But to give up inductive inference entirely, which neither realists nor anti-realists wish to do, is just to stop doing science at all.

There is, in fact, a range of possibilities here where one might be tempted to draw a line. Consider the following:

- I Evidence E is true.
- II Theory T is empirically adequate.
- III The entities T posits exist.
- IV T is true.

They are ordered in terms of decreasing probability, given evidence E. An inductive sceptic will, of course, accept I given E, but will go no further. Van Fraassen will accept the likes of II, given E, but resists III and IV. The niche between II and IV is interesting, though it is not a common position. Nancy Cartwright, for example, believes that there are electrons, but that the electron theory is false. (Her half realist/half anti-realist view is partly revealed in the title of her new book, *How the Laws of Physics Lie*.¹¹)* The full-blooded realist is prepared in principle to accept IV. All of this makes it clear that there are anti-realist positions between full re-

* See Nancy Cartwright, "Do the Laws of Physics State the Facts?" in chapter 7.

alism and inductive scepticism. The fight, contrary to Laudan, is not over the legitimacy of induction.

■ | The Truth Matters (A Little)

It is now time to take stock. By explaining success, remember, there are three things to be accounted for: (1) the fact that theories organize, unify, and generally account for a wide variety of phenomena. (2) Theories are getting better and better at this; they are progressing, and (3) a statistically significant number of their novel predictions are true. It is now time to stand back and see where we have gotten to.

(A) Well, for one thing, this attempt to account for the success of science is not just the problem of induction. So there is some hope of coming up with an answer. Realist explanations of success may well beg the question against that age old problem, but then we all (including the anti-realist) do that all the time. Induction, in principle, is not what is at issue here; rather it is a *particular* inductive inference that is being debated.

(B) The Darwinian account has no answer at all for (3), that is, for the fact that theories make novel predictions which are found to be true. It has an explanation of (1), the significant degree of empirical adequacy, and (2), the increasing degree of empirical adequacy over time, but it can explain these only by linking rational theory choice analytically to success. Since this is methodologically implausible, even the explanations of (1) and (2) are thus not acceptable. So the Darwinian account completely collapses; and quite possibly the hopes of the anti-realist to explain the success of science without appeals to *truth* also collapse. So the realist is now without a plausible rival.

(C) Let us turn now to the realist's account of things. Explaining the second aspect of success (i.e., that the succession of theories is getting better and better at accounting for the phenomena) is probably the most popular approach. Leplin thinks it is the most promising¹² and Newton-Smith, as we saw earlier, builds his doctrine of verisimilitude around it. Actually, it may be the least promising. The realist explanation of this sense of the success of science quite explicitly needs a theory of truthlikeness, that is, it is entirely dependent on the in principle existence of some theory of verisimilitude. However, none is available. I criticized Newton-Smith above and other versions of the doctrine have not gone unscathed either. The historical record makes the prospects for one look rather dim. The whole notion of verisimilitude may have to go the way of, say, 'simplicity'.

(D) The third sense of success (i.e., the making of novel predictions) seems also to be promising. Predictions about the future which turn out

to be true are not just lucky guesses on the realist's account. These predictions are deduced from the truth, says the realist, so it is no wonder the 'guesses' panned out. There is no rival explanation for this; the Darwinian explanation didn't even try to account for it. In Laudan's very detailed attack on convergent realism¹³ there is very little mention of this sense of success. So it remains, it seems to me, something the realist might point to as a genuine accomplishment, something the anti-realist fails to do justice to. But how strong is this? How much support does this give to the realist? Unfortunately many theories now thought to be false made true novel predictions. Ptolemaic astronomy, for instance, predicted eclipses fairly accurately. And Fresnel rather surprisingly got right his prediction of a bright spot in the middle of a shadow cast by a disk. Being true is not necessary for making successful predictions.

(E) It is the first sense of success that seems the least promising for the realist; and the reason is obvious. *Ad hoc* moves can always be employed to do justice to the known phenomena. It is very easy to suspect that this could be going on here. Moreover, the historical record is full of theories which were successful but false, or theories which were unsuccessful but (by our present lights) true. Truth is neither a necessary nor a sufficient condition for success in the first sense, so its explanatory prospects seem dim. This assessment, though it seems obvious, may not be right. In fact, it is with the first sense of success that the realist may have the most hope.

It is hard to say why realist accounts of the success of science have gone wrong. Of course, one answer is that realism itself is wrong. But this is an answer I am loath to accept, so before I do I want to explore at least one different kind of approach to the problem. What realists need, I suggest, is a different style of explanation entirely. I will now try to spell this out, if only briefly, in the balance of this paper. I stress the tentative, exploratory, and sketchy nature of the proposal below; it is intended merely as a beginning.

The last three decades have seen considerable quarreling over the form of a proper explanation. The dominant theory has been the so-called deductive-nomological or covering law model proposed by Hempel. For probabilistic situations there is the so-called inductive-statistical account.* Either way, on Hempel's view, an explanation is an argument. Given the explanans, the explanandum is shown to have been expected. (In the deductive case it is certainly expected and in the inductive case the explanandum is expected with high probability.) In short, an explanation is a sufficient or almost sufficient condition for what is being explained.

Here lies the difficulty. The preceding considerations show that truth is neither a necessary nor a sufficient condition for the success of science.

* See Carl Hempel, "Two Basic Types of Scientific Explanation" and "Inductive-Statistical Explanation" in chapter 6.

It does not meet the Hempelian conditions at all. And since it is not even close to being sufficient we cannot subsume it under the inductive-statistical version of the covering law model either. But the idea that it might have something to do with statistical considerations is, I think, an idea worth exploring.

Wesley Salmon has proposed¹⁴ an account of explanation which rivals the covering law account of Hempel. An explanation is not an argument for a conclusion, it is instead the marshalling of the statistically relevant facts which have a bearing on the outcome. His view was introduced to cope with examples such as this: 'Why does Jones have paresis?' Explanation: 'Because he had syphilis.' This seems intuitively like a good explanation, yet the outcome, Jones's paresis, is not likely at all. The chances of getting paresis are very small with syphilis, but larger than they would be without it. Having syphilis, says Salmon, is statistically relevant; that is why it explains Jones's paresis. (According to Salmon, A is statistically relevant to B if and only if $\text{Prob}(B/A) \neq \text{Prob}(B)$.)¹⁵

We know that false premisses can yield true conclusions, so the truth is not (logically speaking) necessary for success. The reason truth is not sufficient for success is because of the presence of auxiliary assumptions which are also at work in any explanation. Though truth is neither sufficient nor necessary for success, it is, I shall say following Salmon, statistically relevant. The truth matters to the outcome, though it only matters a little.

Salmon's statistical relevance model is not the only challenger to the Hempelian account. Some philosophers of biology¹⁶ and other philosophers of history¹⁷ have advocated a narrative style of explanation. An event or condition is explained by telling a story in which it is embedded. In this way the explanandum is said to be rendered 'intelligible'. It is often claimed that Darwinian evolution, for instance, is unable to satisfy the Hempelian form, but that it is explanatory nevertheless. It provides neither necessary nor sufficient conditions, but it succeeds in some sense or other in explaining things.

Consider a brief example: Why does the giraffe have a long neck? Explanation: The ancestors of the modern giraffe fed on trees, and those with long necks were able to reach more when food was scarce (such as in the occasional drought which may have occurred.) There might have been survival value in having a long neck, so there was, consequently, differential selection in its favour. Is this meant by the evolutionist to be true? Not with any degree of confidence. It is only meant to be an evolutionary *possibility*, one of the many courses that nature *might* have taken.

Narrative explanations are very similar to statistical relevance explanations. Neither provide necessary or sufficient conditions for what is being explained. What both do, however, is provide something which is relevant to the outcome. There is also a difference between them. The statistically relevant information in, for example, the Jones paresis case is

the *known* fact that Jones had syphilis. In typical narrative explanations the statistically relevant fact in the explanation is not known to be true. It is conjectured. (We might, if we wanted to coin a barbaric phrase, call the combined view 'the hypothetico-statistical relevance model of explanation'.)

My suggestion now is simply this: The realist has an explanation for the success of science: Truth is the explanation and the style of the explanation is narrative. The truth is not known to obtain; it is hypothetical. But even if it did obtain, success would not automatically follow. The presence of the truth does make a difference, however; truth is statistically relevant.

The Hempel model of explanation was tied to confirmation. By deducing the data from the theory the theory explained the data and in turn the data confirmed the theory. Alas, this is not the case here. Saying that a theory is true does not lead to any testable predictions over and above those already made by saying that the theory is empirically adequate. There is no additional predictive power to this sort of narrative explanation. But even though predictive power is lost, this does not lead to the demise of the claim to have explanatory power. We cannot predict why a radio-active atom decays at the precise moment that it does; but after it happens we can explain it. The fact that the quantum theory can give such *post hoc* explanations does count in its favour, though only very little. The explanatory power of truth is similar.

It may sound as if the power of truth has become pretty vacuous. We all know the story about a scientist who was asked by an assistant how the theory explained some puzzling data. 'That's easy,' the scientist said, and proceeded to give an account. Later the assistant rushed anxiously back in and reported that the actual data were quite contrary to the earlier report. 'Oh', said the scientist, 'That's even easier to explain.'

Things are not quite that bad. But then, the explanatory power that truth does have in accounting for the success of science is not of the sort to make us believe in realism. For that we will need other considerations, such as, say, the (*a priori*) common cause argument of Salmon,¹⁸ or Hacking's intervening arguments.^{19*} We cannot rely on the success of science as characterized above.

In most explanations there is a connection to justification. That is why Popper does not want truth to explain success. But there are also

* For Salmon's appeal to the principle of the common cause in defense of scientific realism, see Wesley C. Salmon, *Scientific Explanation and the Causal Structure of the World* (Princeton, N.J.: Princeton University Press, 1984), ch. 8. "The common cause principle states, roughly, that when apparent coincidences are too improbable to be attributed to chance, they can be explained by reference to a common causal antecedent" (Salmon, 158). Hacking's intervening argument for realism about theoretical entities can be found in "Experimentation and Scientific Realism," which is the next reading in this chapter.

explanations which are not linked to justification and that, I think, is what is going on here. We show how, given realism, the success of science is possible, why it is not a miracle. But the style of the explanation does not let us infer its correctness.²⁰

■ | Notes

1. *Objective Knowledge*, Oxford: Oxford University Press, 1973, p. 23.
2. *Ibid.*, p. 204.
3. Smart, *Between Science and Philosophy*, New York: Random House, 1968, p. 39.
4. B. van Fraassen, *The Scientific Image*, Oxford: Oxford University Press, 1980, p. 39f. [Excerpted in this chapter, 1064–87]
5. H. Putnam, *Philosophical Papers*, vol. I, Cambridge: Cambridge University Press, 1975, p. 73.
6. W. Newton-Smith, *The Rationality of Science*, London: Routledge and Kegan Paul, 1981, p. 196.
7. *Ibid.*, p. 198.
8. *Ibid.*, p. 204.
9. Laudan, 'A Confutation of Convergent Realism', *Philosophy of Science*, 1981, p. 45. [Excerpted in this chapter, 1114–35]
10. See Laudan, 'Sources of Modern Methodology', reprinted in his *Science and Hypothesis*, Dordrecht: Reidel, 1981.
11. Oxford: Oxford University Press, 1983. The idea is that laws and theories generally are false, but the things they talk about, electrons, genes, etc., are quite real and certainly exist, according to Cartwright.
12. J. Leplin, 'The Historical Objection to Scientific Realism', Asquith and Giere (eds.) *PSA 1980*, vol. I.
13. Laudan, 'A Confutation of Convergent Realism', *loc. cit.*
14. See his 'Statistical Explanation' reprinted in Salmon (ed.) *Statistical Explanation and Statistical Relevance*, Pittsburgh: Pittsburgh University Press, 1971.
15. There are problems with this account; see for example the relevant discussion by Cartwright in *How the Laws of Physics Lie*, *loc. cit.* Salmon has further fine tuned his view in "Why Ask 'Why?'", *Proceedings of the American Philosophical Association*, 1978.
16. For example, Goudge, *The Ascent of Life*, London: George Allen and Unwin, 1961.
17. For example, Dray, *Philosophy of History*, Engelwood Cliffs: Prentice-Hall, 1964.
18. "Why Ask 'Why?'", *loc. cit.*

19. *Representing and Intervening*, Cambridge: Cambridge University Press, 1983.

20. This paper was presented to the philosophy of science conference in Dubrovnik, Yugoslavia, April 1984. It has benefited from the comments of those present. Research supported, in part, by a grant from S.S.H.R.C.

Experimentation and Scientific Realism

Experimental physics provides the strongest evidence for scientific realism. Entities that in principle cannot be observed are regularly manipulated to produce new phenomena and to investigate other aspects of nature. They are tools, instruments not for thinking but for doing.

The philosopher's standard "theoretical entity" is the electron. I shall illustrate how electrons have become experimental entities, or experimenter's entities. In the early stages of our discovery of an entity, we may test hypotheses about it. Then it is merely an hypothetical entity. Much later, if we come to understand some of its causal powers and to use it to build devices that achieve well understood effects in other parts of nature, then it assumes quite a different status.

Discussions about scientific realism or anti-realism usually talk about theories, explanation and prediction. Debates at that level are necessarily inconclusive. Only at the level of experimental practice is scientific realism unavoidable. But this realism is not about theories and truth. The experimentalist need only be a realist about the entities used as tools.

■ | A Plea for Experiments

No field in the philosophy of science is more systematically neglected than experiment. Our grade school teachers may have told us that scientific method is experimental method, but histories of science have become histories of theory. Experiments, the philosophers say, are of value only when they test theory. Experimental work, they imply, has no life of its own. So we lack even a terminology to describe the many varied roles of experiment. Nor has this one-sidedness done theory any good, for radically different types of theory are used to think about the same physical phe-

nomenon (e.g., the magneto-optical effect). The philosophers of theory have not noticed this and so misreport even theoretical inquiry.¹

Different sciences at different times exhibit different relationships between "theory" and "experiment." One chief role of experiment is the creation of phenomena. Experimenters bring into being phenomena that do not naturally exist in a pure state. These phenomena are the touchstones of physics, the keys to nature and the source of much modern technology. Many are what physicists after the 1870s began to call "effects": the photo-electric effect, the Compton effect, and so forth. A recent high-energy extension of the creation of phenomena is the creation of "events," to use the jargon of the trade. Most of the phenomena, effects and events created by the experimenter are like plutonium: they do not exist in nature except possibly on vanishingly rare occasions.²

In this paper I leave aside questions of methodology, history, taxonomy and the purpose of experiment in natural science. I turn to the purely philosophical issue of scientific realism. Call it simply "realism" for short. There are two basic kinds: realism about entities and realism about theories. There is no agreement on the precise definition of either. Realism about theories says we try to form true theories about the world, about the inner constitution of matter and about the outer reaches of space. This realism gets its bite from optimism; we think we can do well in this project, and have already had partial success.

Realism about entities—and I include processes, states, waves, currents, interactions, fields, black holes and the like among entities—asserts the existence of at least some of the entities that are the stock in trade of physics.³

The two realisms may seem identical. If you believe a theory, do you not believe in the existence of the entities it speaks about? If you believe in some entities, must you not describe them in some theoretical way that you accept? This seeming identity is illusory. The vast majority of experimental physicists are realists about entities without a commitment to realism about theories. The experimenter is convinced of the existence of plenty of "inferred" and "unobservable" entities. But no one in the lab believes in the literal truth of present theories about those entities. Although various properties are confidently ascribed to electrons, most of these properties can be embedded in plenty of different inconsistent theories about which the experimenter is agnostic. Even people working on adjacent parts of the same large experiment will use different and mutually incompatible accounts of what an electron is. That is because different parts of the experiment will make different uses of electrons, and the models that are useful for making calculations about one use may be completely haywire for another use.

Do I describe a merely sociological fact about experimentalists? It is not surprising, it will be said, that these good practical people are realists. They need that for their own self-esteem. But the self-vindicating realism

of experimenters shows nothing about what actually exists in the world. In reply I repeat the distinction between realism about entities and realism about theories and models. Anti-realism about models is perfectly coherent. Many research workers may in fact hope that their theories and even their mathematical models “aim at the truth,” but they seldom suppose that any particular model is more than adequate for a purpose. By and large most experimenters seem to be instrumentalists about the models they use. The models are products of the intellect, tools for thinking and calculating. They are essential for writing up grant proposals to obtain further funding. They are rules of thumb used to get things done. Some experimenters are instrumentalists about theories and models, while some are not. That is a sociological fact. But experimenters are realists about the entities that they use in order to investigate other hypotheses or hypothetical entities. That is not a sociological fact. Their enterprise would be incoherent without it. But their enterprise is not incoherent. It persistently creates new phenomena that become regular technology. My task is to show that realism about entities is a necessary condition for the coherence of most experimentation in natural science.

■ | Our Debt to Hilary Putnam

It was once the accepted wisdom that a word like “electron” gets its meaning from its place in a network of sentences that state theoretical laws. Hence arose the infamous problems of incommensurability and theory change. For if a theory is modified, how could a word like “electron” retain its previous meaning? How could different theories about electrons be compared, since the very word “electron” would differ in meaning from theory to theory?

Putnam saves us from such questions by inventing a referential model of meaning. He says that meaning is a vector, refreshingly like a dictionary entry. First comes the syntactic marker (part of speech). Next the semantic marker (general category of thing signified by the word). Then the stereotype (clichés about the natural kind, standard examples of its use and present day associations. The stereotype is subject to change as opinions about the kind are modified). Finally there is the actual reference of the word, the very stuff, or thing, it denotes if it denotes anything. (Evidently dictionaries cannot include this in their entry, but pictorial dictionaries do their best by inserting illustrations whenever possible.)⁴

Putnam thought we can often guess at entities that we do not literally point to. Our initial guesses may be jejune or inept, and not every naming of an invisible thing or stuff pans out. But when it does, and we frame better and better ideas, then Putnam says that although the stereotype changes, we refer to the same kind of thing or stuff all along. We and

Dalton alike spoke about the same stuff when we spoke of (inorganic) acids. J. J. Thomson, Lorentz, Bohr and Millikan were, with their different theories and observations, speculating about the same kind of thing, the electron.

There is plenty of unimportant vagueness about when an entity has been successfully “dubbed,” as Putnam puts it. “Electron” is the name suggested by G. Johnstone Stoney in 1891 as the name for a natural unit of electricity. He had drawn attention to this unit in 1874. The name was then applied in 1897 by J. J. Thomson to the subatomic particles of negative charge of which cathode rays consist. Was Johnstone Stoney referring to the electron? Putnam’s account does not require an unequivocal answer. Standard physics books say that Thomson discovered the electron. For once I might back theory and say Lorentz beat him to it. What Thomson did was to measure the electron. He showed its mass is $\frac{1}{1800}$ that of hydrogen. Hence it is natural to say that Lorentz merely postulated the particle of negative charge, while Thomson, determining its mass, showed that there is some such real stuff beaming off a hot cathode.

The stereotype of the electron has regularly changed, and we have at least two largely incompatible stereotypes, the electron as cloud and the electron as particle. One fundamental enrichment of the idea came in the 1920s. Electrons, it was found, have angular momentum, or “spin.” Experimental work by Stern and Gerlach first indicated this, and then Goudsmit and Uhlenbeck provided the theoretical understanding of it in 1925. Whatever we think about Johnstone Stoney, others—Lorentz, Bohr, Thomson and Goudsmit—were all finding out more about the same kind of thing, the electron.

We need not accept the fine points of Putnam’s account of reference in order to thank him for providing a new way to talk about meaning. Serious discussions of inferred entities need no longer lock us into pseudo-problems of incommensurability and theory change. Twenty-five years ago the experimenter who believed that electrons exist, without giving much credence to any set of laws about electrons, would have been dismissed as philosophically incoherent. We now realize it was the philosophy that was wrong, not the experimenter. My own relationship to Putnam’s account of meaning is like the experimenter’s relationship to a theory. I don’t literally believe Putnam, but I am happy to employ his account as an alternative to the unpalatable account in fashion some time ago. . . .

■ | Interfering

Francis Bacon, the first and almost last philosopher of experiments, knew it well: the experimenter sets out “to twist the lion’s tail.” Experimentation is interference in the course of nature; “nature under constraint and vexed;

that is to say, when by art and the hand of man she is forced out of her natural state, and squeezed and moulded.”⁵ The experimenter is convinced of the reality of entities some of whose causal properties are sufficiently well understood that they can be used to interfere *elsewhere* in nature. One is impressed by entities that one can use to test conjectures about other more hypothetical entities. In my example, one is sure of the electrons that are used to investigate weak neutral currents and neutral bosons. This should not be news, for why else are we (non-sceptics) sure of the reality of even macroscopic objects, but because of what we do with them, what we do to them, and what they do to us?

Interference and intervention are the the stuff of reality. This is true, for example, at the borderline of observability. Too often philosophers imagine that microscopes carry conviction because they help us see better. But that is only part of the story. On the contrary, what counts is what we can do to a specimen under a microscope, and what we can see ourselves doing. We stain the specimen, slice it, inject it, irradiate it, fix it. We examine it using different kinds of microscopes that employ optical systems that rely on almost totally unrelated facts about light. Microscopes carry conviction because of the great array of interactions and interferences that are possible. When we see something that turns out not to be stable under such play, we call it an artefact and say it is not real.⁶

Likewise, as we move down in scale to the truly un-seeable, it is our power to use unobservable entities that make us believe they are there. Yet I blush over these words “see” and “observe.” John Dewey would have said that a fascination with seeing-with-the-naked-eye is part of the Spectator Theory of Knowledge that has bedeviled philosophy from earliest times. But I don’t think Plato or Locke or anyone before the nineteenth century was as obsessed with the sheer opacity of objects as we have been since. My own obsession with a technology that manipulates objects is, of course a twentieth-century counterpart to positivism and phenomenology. Their proper rebuttal is not a restriction to a narrower domain of reality, namely to what can be positivistically “seen” (with the eye), but an extension to other modes by which people can extend their consciousness.

■ | Making

Even if experimenters are realists about entities, it does not follow that they are right. Perhaps it is a matter of psychology: the very skills that make for a great experimenter go with a certain cast of mind that objectifies whatever it thinks about. Yet this will not do. The experimenter cheerfully regards neutral bosons as merely hypothetical entities, while electrons are real. What is the difference?

There are an enormous number of ways to make instruments that rely

on the causal properties of electrons in order to produce desired effects of unsurpassed precision. I shall illustrate this. The argument—it could be called the experimental argument for realism—is not that we infer the reality of electrons from our success. We do not make the instruments and then infer the reality of the electrons, as when we test a hypothesis, and then believe it because it passed the test. That gets the time-order wrong. By now we design apparatus relying on a modest number of home truths about electrons to produce some other phenomenon that we wish to investigate.

That may sound as if we believe in the electrons because we predict how our apparatus will behave. That too is misleading. We have a number of general ideas about how to prepare polarized electrons, say. We spend a lot of time building prototypes that don't work. We get rid of innumerable bugs. Often we have to give up and try another approach. Debugging is not a matter of theoretically explaining or predicting what is going wrong. It is partly a matter of getting rid of "noise" in the apparatus. "Noise" often means all the events that are not understood by any theory. The instrument must be able to isolate, physically, the properties of the entities that we wish to use, and damp down all the other effects that might get in our way. *We are completely convinced of the reality of electrons when we regularly set out to build—and often enough succeed in building—new kinds of devices that use various well understood causal properties of electrons to interfere in other more hypothetical parts of nature.*

It is not possible to grasp this without an example. Familiar historical examples have usually become encrusted by false theory-oriented philosophy or history. So I shall take something new. This is a polarizing electron gun whose acronym is PEGGY II. In 1978 it was used in a fundamental experiment that attracted attention even in *The New York Times*. In the next section I describe the point of making PEGGY II. So I have to tell some new physics. You can omit this and read only the engineering section that follows. Yet it must be of interest to know the rather easy-to-understand significance of the main experimental results, namely, (1) parity is not conserved in scattering of polarized electrons from deuterium, and (2) more generally, parity is violated in weak neutral current interactions.⁷

■ | Methodological Remark

In the following section I retail a little current physics; in the section after that I describe how a machine has been made. It is the latter that matters to my case, not the former. Importantly, even if present quantum electrodynamics turns out to need radical revision, the machine, called PEGGY II, will still work. I am concerned with how it was made to work,

and why. I shall sketch far more sheer engineering than is seen in philosophy papers. My reason is that the engineering is incoherent unless electrons are taken for granted. One cannot say this by merely reporting, "Oh, they made an electron gun for shooting polarized electrons." An immense practical knowledge of how to manipulate electrons, of what sorts of things they will do reliably and how they tend to misbehave—that is the kind of knowledge which grounds the experimenter's realism about electrons. You cannot grasp this kind of knowledge in the abstract, for it is practical knowledge. So I must painfully introduce the reader to some laboratory physics. Luckily it is a lot of fun.

■ | Parity and Weak Neutral Currents

There are four fundamental forces in nature, not necessarily distinct. Gravity and electromagnetism are familiar. Then there are the strong and weak forces, the fulfillment of Newton's program, in the *Optics*, which taught that all nature would be understood by the interaction of particles with various forces that were effective in attraction or repulsion over various different distances (i.e., with different rates of extinction).

Strong forces are 100 times stronger than electromagnetism but act only for a miniscule distance, at most the diameter of a proton. Strong forces act on "hadrons," which include protons, neutrons, and more recent particles, but not electrons or any other members of the class of particles called "leptons."

The weak forces are only $1/10,000$ times as strong as electromagnetism, and act over a distance $1/100$ times smaller than strong forces. But they act on both hadrons and leptons, including electrons. The most familiar example of a weak force may be radioactivity.

The theory that motivates such speculation is quantum electrodynamics. It is incredibly successful, yielding many predictions better than one part in a million, a miracle in experimental physics. It applies over distances ranging from diameters of the earth to $1/100$ the diameter of the proton. This theory supposes that all the forces are "carried" by some sort of particle. Photons do the job in electromagnetism. We hypothesize "gravitons" for gravity.

In the case of interactions involving weak forces, there are charged currents. We postulate that particles called bosons carry these weak forces.⁹ For charged currents, the bosons may be positive or negative. In the 1970s there arose the possibility that there could be weak "neutral" currents in which no charge is carried or exchanged. By sheer analogy with the vindicated parts of quantum electrodynamics, neutral bosons were postulated as the carriers in weak interactions.

The most famous discovery of recent high energy physics is the failure

of the conservation of parity. Contrary to the expectations of many physicists and philosophers, including Kant,⁹ nature makes an absolute distinction between right-handedness and left-handedness. Apparently this happens only in weak interactions.*

What we mean by right- or left-handed in nature has an element of convention. I remarked that electrons have spin. Imagine your right hand wrapped around a spinning particle with the fingers pointing in the direction of spin. Then your thumb is said to point in the direction of the spin vector. If such particles are traveling in a beam, consider the relation between the spin vector and the beam. If all the particles have their spin vector in the same direction as the beam, they have right-handed (linear) polarization, while if the spin vector is opposite to the beam direction, they have left-handed (linear) polarization.

The original discovery of parity violation showed that one kind of product of a particle decay, a so-called *muon neutrino*, exists only in left-handed polarization and never in right-handed polarization.

Parity violations have been found for weak *charged* interactions. What about weak *neutral* currents? The remarkable Weinberg-Salam model for the four kinds of force was proposed independently by Steven Weinberg in 1967 and A[bdu] Salam in 1968. It implies a minute violation of parity in weak neutral interactions. Given that the model is sheer speculation, its success has been amazing, even awe inspiring. So it seemed worthwhile to try out the predicted failure of parity for weak neutral interactions. That would teach us more about those weak forces that act over so minute a distance.

The prediction is: Slightly more left-handed polarized electrons hitting certain targets will scatter, than right-handed electrons. Slightly more! The difference in relative frequency of the two kinds of scattering is one part in 10,000, comparable to a difference in probability between 0.50005 and 0.49995. Suppose one used the standard equipment available at the Stanford Linear Accelerator in the early 1970s, generating 120 pulses per second, each pulse providing one electron event. Then you would have to run the entire SLAC [Stanford Linear Accelerator Center] beam for 27 years in order to detect so small a difference in relative frequency. Considering that one uses the same beam for lots of experiments simultaneously, by letting different experiments use different pulses, and considering

* Kant did not deny that there is a real, intrinsic difference between right-handed and left-handed objects. Hacking is referring to the belief, held universally prior to 1956, that the laws of nature are indifferent to the left-right distinction: for any process that is physically possible, so too is its mirror image. It is this belief that parity-violation experiments proved false. See Martin Gardner, *The New Ambidextrous Universe*, 3d rev. ed. (New York: W. H. Freeman and Company, 1990), ch. 22, and James Van Cleve and Robert E. Frederick, eds., *The Philosophy of Right and Left: Incongruent Counterparts and the Nature of Space* (Dordrecht, Netherlands: Kluwer, 1991).

that no equipment remains stable for even a month, let alone 27 years, such an experiment is impossible. You need enormously more electrons coming off in each pulse. We need between 1000 and 10,000 more electrons per pulse than was once possible. The first attempt used an instrument now called PEGGY I. It had, in essence, a high-class version of J. J. Thomson's hot cathode. Some lithium was heated and electrons were boiled off. PEGGY II uses quite different principles.

■ | PEGGY II

The basic idea began when C. Y. Prescott noticed, (by "chance"!) an article in an optics magazine about a crystalline substance called gallium arsenide. GaAs has a number of curious properties that make it important in laser technology. One of its quirks is that when it is struck by circularly polarized light of the right frequencies, it emits a lot of linearly polarized electrons. There is a good rough and ready quantum understanding of why this happens, and why half the emitted electrons will be polarized, $\frac{3}{4}$ polarized in one direction and $\frac{1}{4}$ polarized in the other.

PEGGY II uses this fact, plus the fact that GaAs emits lots of electrons due to features of its crystal structure. Then comes some engineering. It takes work to liberate an electron from a surface. We know that painting a surface with the right substance helps. In this case, a thin layer of cesium and oxygen is applied to the crystal. Moreover the less air pressure around the crystal, the more electrons will escape for a given amount of work. So the bombardment takes place in a good vacuum at the temperature of liquid nitrogen.

We need the right source of light. A laser with bursts of red light (7100 Ångströms) is trained on the crystal. The light first goes through an ordinary polarizer, a very old-fashioned prism of calcite, or Iceland spar.¹⁰ This gives linearly polarized light. We want circularly polarized light to hit the crystal. The polarized laser beam now goes through a cunning modern device, called a Pockels cell. It electrically turns linearly polarized photons into circularly polarized ones. Being electric, it acts as a very fast switch. The direction of circular polarization depends on the direction of current in the cell. Hence the direction of polarization can be varied randomly. This is important, for we are trying to detect a minute asymmetry between right- and left-handed polarization. Randomizing helps us guard against any systematic "drift" in the equipment.¹¹ The randomization is generated by a radioactive decay device, and a computer records the direction of polarization for each pulse.

A circularly polarized pulse hits the GaAs crystal, resulting in a pulse of linearly polarized electrons. A beam of such pulses is maneuvered by magnets into the accelerator for the next bit of the experiment. It passes

through a device that checks on a proportion of polarization along the way. The remainder of the experiment requires other devices and detectors of comparable ingenuity, but let us stop at PEGGY II.

■ | Bugs

Short descriptions make it all sound too easy, so let us pause to reflect on debugging. Many of the bugs are never understood. They are eliminated by trial and error. Let us illustrate three different kinds: (1) The essential technical limitations that in the end have to be factored into the analysis of error. (2) Simpler mechanical defects you never think of until they are forced on you. (3) Hunches about what might go wrong.

1. Laser beams are not as constant as science fiction teaches, and there is always an irremediable amount of “jitter” in the beam over any stretch of time.

2. At a more humdrum level the electrons from the GaAs crystal are back-scattered and go back along the same channel as the laser beam used to hit the crystal. Most of them are then deflected magnetically. But some get reflected from the laser apparatus and get back into the system. So you have to eliminate these new ambient electrons. This is done by crude mechanical means, making them focus just off the crystal and so wander away.

3. Good experimenters guard against the absurd. Suppose that dust particles on an experimental surface lie down flat when a polarized pulse hits them, and then stand on their heads when hit by a pulse polarized in the opposite direction. Might that have a systematic effect, given that we are detecting a minute asymmetry? One of the team thought of this in the middle of the night and came down next morning frantically using antidust spray. They kept that up for a month, just in case.¹²

■ | Results

Some 10^{11} events were needed to obtain a result that could be recognized above systematic and statistical error. Although the idea of systematic error presents interesting conceptual problems, it seems to be unknown to philosophers. There were systematic uncertainties in the detection of right- and left-handed polarization, there was some jitter, and there were other problems about the parameters of the two kinds of beam. These errors were analyzed and linearly added to the statistical error. To a student of statistical inference this is real seat-of-the-pants analysis with no rationale whatsoever. Be that as it may, thanks to PEGGY II the number of events was big enough to give a result that convinced the entire physics com-

munity.¹³ Left-handed polarized electrons were scattered from deuterium slightly more frequently than right-handed electrons. This was the first convincing example of parity-violation in a weak neutral current interaction.*

■ | Comment

The making of PEGGY II was fairly non-theoretical. Nobody worked out in advance the polarizing properties of GaAs—that was found by a chance encounter with an unrelated experimental investigation. Although elementary quantum theory of crystals explains the polarization effect, it does not explain the properties of the actual crystal used. No one has been able to get a real crystal to polarize more than 37 percent of the electrons, although in principle 50 percent should be polarized.

Likewise although we have a general picture of why layers of cesium and oxygen will “produce negative electron affinity,” i.e., make it easier for electrons to escape, we have no quantitative understanding of why this increases efficiency to a score of 37 percent.

Nor was there any guarantee that the bits and pieces would fit together. To give an even more current illustration, future experimental work, briefly described later in this paper, makes us want even more electrons per pulse than PEGGY II could give. When the parity experiment was reported in *The New York Times*, a group at Bell Laboratories read the newspaper and saw what was going on. They had been constructing a crystal lattice for totally unrelated purposes. It uses layers of GaAs and a related aluminum compound. The structure of this lattice leads one to expect that virtually all the electrons emitted would be polarized. So we might be able to double the efficiency of PEGGY II. But at present (July 1981) that nice idea has problems. The new lattice should also be coated in work-reducing paint. But the cesium-oxygen compound is applied at high temperature. Then the aluminum tends to ooze into the neighboring layer of GaAs, and the pretty artificial lattice becomes a bit uneven, limiting its fine polarized-electron-emitting properties. So perhaps this will never work.¹⁴ The group are simultaneously reviving a souped up new thermionic cathode to try to get more electrons. Maybe PEGGY II would have shared the same fate, never working, and thermionic devices would have stolen the show.

Note, incidentally, that the Bell people did not need to know a lot of

* For competing accounts of why physicists found the SLAC experiment so convincing, see Andrew Pickering, *Constructing Quarks* (Chicago: University of Chicago Press, 1984), and Allan Franklin, *Experiment, Right or Wrong* (Cambridge: Cambridge University Press, 1990).

weak neutral current theory to send along their sample lattice. They just read *The New York Times*.

■ | Moral

Once upon a time it made good sense to doubt that there are electrons. Even after Millikan had measured the charge on the electron, doubt made sense. Perhaps Millikan was engaging in “inference to the best explanation.” The charges on his carefully selected oil drops were all small integral multiples of a least charge. He inferred that this is the real least charge in nature, and hence it is the charge on the electron, and hence there are electrons, particles of least charge. In Millikan’s day most (but not all) physicists did become increasingly convinced by one or more theories about the electron. However it is always admissible, at least for philosophers, to treat inferences to the best explanation in a purely instrumental way, without any commitment to the existence of entities used in the explanation.¹⁵ But it is now seventy years after Millikan, and we no longer have to infer from explanatory success. Prescott et al., don’t explain phenomena with electrons. They know a great deal about how to use them.

The group of experimenters do not know what electrons are, exactly. Inevitably they think in terms of particles. There is also a cloud picture of an electron which helps us think of complex wave functions of electrons in a bound state. The angular momentum and spin vector of a cloud make little sense outside a mathematical formalism. A beam of polarized clouds is fantasy so no experimenter uses that model—not because of doubting its truth, but because other models help more with the calculations. Nobody thinks that electrons “really” are just little spinning orbs about which you could, with a small enough hand, wrap the fingers and find the direction of spin along the thumb. There is instead a family of causal properties in terms of which gifted experimenters describe and deploy electrons in order to investigate something else, e.g., weak neutral currents and neutral bosons. We know an enormous amount about the behavior of electrons. We also know what does not matter to electrons. Thus we know that bending a polarized electron beam in magnetic coils does not affect polarization in any significant way. We have hunches, too strong to ignore although too trivial to test independently: e.g., dust might dance under changes of directions of polarization. Those hunches are based on a hard-won sense of the kinds of things electrons are. It does not matter at all to this hunch whether electrons are clouds or particles.

The experimentalist does not believe in electrons because, in the words retrieved from mediaeval science by Duhem, they “save the phenomena.” On the contrary, we believe in them because we use them to

create new phenomena, such as the phenomenon of parity violation in weak neutral current interactions.

■ | When Hypothetical Entities Become Real

Note the complete contrast between electrons and neutral bosons. Nobody can yet manipulate a bunch of neutral bosons, if there are any. Even weak neutral currents are only just emerging from the mists of hypothesis. By 1980 a sufficient range of convincing experiments had made them the object of investigation. When might they lose their hypothetical status and become commonplace reality like electrons? When we use them to investigate something else.

I mentioned the desire to make a better gun than PEGGY II. Why? Because we now “know” that parity is violated in weak neutral interactions. Perhaps by an even more grotesque statistical analysis than that involved in the parity experiment, we can isolate just the weak interactions. That is, we have a lot of interactions, including say electromagnetic ones. We can censor these in various ways, but we can also statistically pick out a class of weak interactions as precisely those where parity is not conserved. This would possibly give us a road to quite deep investigations of matter and anti-matter. To do the statistics one needs even more electrons per pulse than PEGGY II could hope to generate. If such a project were to succeed, we should be beginning to use weak neutral currents as a manipulable tool for looking at something else. The next step towards a realism about such currents would have been made.

The message is general and could be extracted from almost any branch of physics. Dudley Shapere has recently used “observation” of the sun’s hot core to illustrate how physicists employ the concept of observation. They collect neutrinos from the sun in an enormous disused underground mine that has been filled with the old cleaning fluid (i.e., carbon tetrachloride). We would know a lot about the inside of the sun if we knew how many solar neutrinos arrive on the earth. So these are captured in the cleaning fluid; a few will form a new radioactive nucleus. The number that do this can be counted. Although the extent of neutrino manipulation is much less than electron manipulation in the PEGGY II experiment, here we are plainly using neutrinos to investigate something else. Yet not many years ago, neutrinos were about as hypothetical as an entity could get. After 1946 it was realized that when mesons disintegrate, giving off, among other things, highly energized electrons, one needed an extra nonionizing particle to conserve momentum and energy. At that time this postulated “neutrino” was thoroughly hypothetical, but now it is routinely used to examine other things.

■ | Changing Times

Although realisms and anti-realisms are part of the philosophy of science well back into Greek prehistory, our present versions mostly descend from debates about atomism at the end of the nineteenth century. Anti-realism about atoms was partly a matter of physics: the energeticists thought energy was at the bottom of everything, not tiny bits of matter. It also was connected with the positivism of Comte, Mach, Pearson and even J. S. Mill. Mill's young associate Alexander Bain states the point in a characteristic way, apt for 1870:

Some hypotheses consist of assumptions as to the minute structure and operations of bodies. From the nature of the case these assumptions can never be proved by direct means. Their merit is their suitability to express phenomena. They are Representative Fictions.¹⁶

"All assertions as to the ultimate structure of the particles of matter," continues Bain, "are and ever must be hypothetical. . . ." The kinetic theory of heat, he says, "serves an important intellectual function." But we cannot hold it to be a true description of the world. It is a Representative Fiction.

Bain was surely right a century ago. Assumptions about the minute structure of matter could not be proved then. The only proof could be indirect, namely that hypotheses seemed to provide some explanation and helped make good predictions. Such inferences need never produce conviction in the philosopher inclined to instrumentalism or some other brand of idealism.

Indeed the situation is quite similar to seventeenth-century epistemology. At that time knowledge was thought of as correct representation. But then one could never get outside the representations to be sure that they corresponded to the world. Every test of a representation is just another representation. "Nothing is so much like an idea as an idea," as Bishop Berkeley had it. To attempt to argue for scientific realism at the level of theory, testing, explanation, predictive success, convergence of theories and so forth is to be locked into a world of representations. No wonder that scientific anti-realism is so permanently in the race. It is a variant on "The Spectator Theory of Knowledge."^{*}

^{*} According to the American philosopher John Dewey (1859–1952), the "spectator theory of knowledge" is the mistake, pervasive among Western philosophers from the time of Plato, of thinking that knowledge involves passively representing the world (where these representations—beliefs—are judged by their correspondence with facts) as opposed to actively constructing conceptual frameworks (where these conceptual frameworks are judged by their instrumental value in predicting experience and guiding our actions).

Scientists, as opposed to philosophers, did in general become realists about atoms by 1910. Michael Gardner, in one of the finest studies of real-life scientific realism, details many of the factors that went into that change in climate of opinion.¹⁷ Despite the changing climate, some variety of instrumentalism or fictionalism remained a strong philosophical alternative in 1910 and in 1930. That is what the history of philosophy teaches us. Its most recent lesson is Bas van Fraassen's *The Scientific Image*, whose "constructive empiricism" is another theory-oriented anti-realism. The lesson is: think about practice, not theory.

Anti-realism about atoms was very sensible when Bain wrote a century ago. Anti-realism about *any* sub-microscopic entities was a sound doctrine in those days. Things are different now. The "direct" proof of electrons and the like is our ability to manipulate them using well understood low-level causal properties. I do not of course claim that "reality" is constituted by human manipulability. We can, however, call something real, in the sense in which it matters to scientific realism, only when we understand quite well what its causal properties are. The best evidence for this kind of understanding is that we can set out, from scratch, to build machines that will work fairly reliably, taking advantage of this or that causal nexus. Hence, engineering, not theorizing, is the proof of scientific realism about entities.¹⁸

■ | Notes

1. C. W. F. Everitt and Ian Hacking, "Which Comes First, Theory or Experiment?" [Unpublished. See chapter 9, "Experiment," of Ian Hacking, *Representing and Intervening* (Cambridge: Cambridge University Press, 1983).]
2. Ian Hacking, "Spekulation, Berechnung und die Erschaffung der Phänomene," in *Versuchungen: Aufsätze zur Philosophie Paul Feyerabends*, (P. Duerr, ed.), Frankfurt, 1981, Bd 2, 126–58.
3. Nancy Cartwright makes a similar distinction in a sequence of papers, including "When Explanation Leads to Inference," [*Philosophical Topics* 13 (1982): 111–21]. She approaches realism from the top, distinguishing theoretical laws (which do not state the facts) from phenomenological laws (which do). She believes in some "theoretical" entities and rejects much theory on the basis of a subtle analysis of modeling in physics. I proceed in the opposite direction, from experimental practice. Both approaches share an interest in real-life physics as opposed to philosophical fantasy science. My own approach owes an enormous amount to Cartwright's parallel developments, which have often preceded my own. My use of the two kinds of realism is a case in point.
4. Hilary Putnam, "How Not To Talk About Meaning," "The Meaning of 'Meaning'," and other papers in the *Mind, Language and Reality*, *Philosophical Papers*, Vol. 2. Cambridge, 1975.
5. Francis Bacon, *The Great Instauration*, in *The Philosophical Works of Francis Bacon* (J. M. Robertson, ed; Ellis and Spedding, Trans.), London, 1905, p. 252.

6. Ian Hacking, "Do We See Through a Microscope?" *Pacific Philosophical Quarterly*, 62, (winter 1981), 305–22.
7. I thank Melissa Franklin, of the Stanford Linear Accelerator, for introducing me to PEGGY II and telling me how it works. She also arranged discussions with members of the PEGGY II group, some of which are mentioned below. The report of experiment E-122 described here is "Parity Non-conservation in Inelastic Electron Scattering," C. Y. Prescott et al., *Physics Letters*. I have relied heavily on the in-house journal, the *SLAC Beam Line*, Report No. 8, October, 1978, "Parity Violation in Polarized Electron Scattering." This was prepared by the in-house science writer Bill Kirk, who is the clearest, most readable popularizer of difficult new experimental physics that I have come across.
8. The odd-sounding bosons are named after the Indian physicist S. N. Bose (1894–1974), also remembered in the name "Bose-Einstein statistics" (which bosons satisfy).
9. But excluding Leibniz, who "knew" there had to be some real, natural difference between right- and left-handedness.
10. Iceland spar is an elegant example of how experimental phenomena persist even while theories about them undergo revolutions. Mariners brought calcite from Iceland to Scandinavia. Erasmus Bartholin experimented with it and wrote about it in 1689. When you look through these beautiful crystals you see double, thanks to the so-called ordinary and extraordinary rays. Calcite is a natural polarizer. It was our entry to polarized light which for 300 years was the chief route to improved theoretical and experimental understanding of light and then electromagnetism. The use of calcite in PEGGY II is a happy reminder of a great tradition.
11. It also turns GaAs, a $\frac{3}{4}$ - $\frac{1}{4}$ left/right hand polarizer, into a 50-50 polarizer.
12. I owe these examples to conversation with Roger Miller of SLAC.
13. The concept of a "convincing experiment" is fundamental. Peter Galison has done important work on this idea, studying European and American experiments on weak neutral currents conducted during the 1970s.
14. I owe this information to Charles Sinclair of SLAC.
15. My attitude to "inference to the best explanation" is one of many learned from Cartwright. See, for example, her paper ["When Explanation Leads to Inference," *loc. cit.*].
16. Alexander Bain, *Logic, Deductive and Inductive*, London and New York, 1870, p. 362.
17. Michael Gardner, "Realism and Instrumentalism in 19th-Century Atomism," *Philosophy of Science* 46, (1979), 1–34.
18. (Added in proof, February, 1983). As indicated in the text, this is a paper of July, 1981, and hence is out of date. For example, neutral bosons are described as purely hypothetical. Their status has changed since CERN announced on Jan. 23, 1983, that a group there had found W, the weak intermediary boson, in proton-antiproton decay at 540 GeV. These experimental issues are further discussed in my . . . book, *Representing and Intervening* (Cambridge, 1983).

Hacking's Experimental Realism

Traditional debates about scientific realism tend to focus on issues concerning scientific representation (broadly speaking) and de-emphasize issues concerning scientific intervention. Questions about the relation between theories and the world, the nature of scientific inference, and the structure of scientific explanations have occupied a central place in the realism debate, while questions about experimentation and technology have not. Ian Hacking's experimental realism attempts to reverse this trend by shifting the defense of realism away from representation to intervention.¹ Experimental realism, according to Hacking, does not require us to believe that our theories are true (or approximately true), nor does its defense depend on inference to the best explanation. For Hacking, the strongest proof for realism is that we can manipulate objects: 'So far as I'm concerned, if you can spray them, then they are real' (ibid., 23).

In this paper I argue (1) that Hacking's argument for experimental realism is, despite his strong denials, another version of the 'success of science' argument; (2) that the experimental realist can only have knowledge about theoretical entities if she assumes that the theories which describe those entities are at least approximately true; and (3) that experimentation is not nearly as theory-free as Hacking maintains. The common thread in these three criticisms is that Hacking does not succeed in shifting the defense of realism away from questions about scientific representation. Experimentation is a form of intervention, but it is intervention strongly guided by representation.

■ | I

Since 'realism' is one of the more abused words in the philosophical lexicon, it is important to understand how Hacking uses it. The word 'real,' according to Hacking, is a 'trouser-word' in that its meaning derives from what it is contrasted with: real ducks are not decoys, real diamonds are not made of cubic zirconium, and so forth.* Real, in its scientific usage, is connected to causation: real entities can causally interact with the world; unreal ones cannot.

Reality has to do with causation and our notions of reality are formed from our ability to change the world. . . . We shall count as real what we can use to intervene in the world to affect something else, or what the world can use to affect us. (ibid., 46)

Given this account of reality, Hacking distinguishes between two types of realism: theory realism and entity realism. Theory realists hold that scientific theories accurately describe the causal structures and properties in the world. They believe that (1) theories are true or false independently of what we believe; (2) truth is determined by the way the world is (i.e., correspondence truth); (3) some theories are approximately true; and (4) theories aim at the truth. Theory anti-realists may deny some or all of these claims. Entity realists, on the other hand, hold only that theoretical entities (or properties, processes, and events) are real; that is, that they are part of the causal structure of the world. Entity realists need not believe that our theories about those entities are true or even close to the truth, according to Hacking. Entity anti-realists claim theoretical entities are useful fictions, logical constructions, or instruments for reasoning (ibid., 27).

Epistemology, not just metaphysics and semantics, also plays a crucial role in Hacking's experimental realism. Hacking, by his own admission, runs together metaphysics and epistemology in his construal of entity realism: the entity realist claims to *know* that some theoretical entities *exist*.

* J. L. Austin coined the term *trouser-word* in his *Sense and Sensibilia* (Oxford: Clarendon Press, 1962), p. 70. See section 7, on the meaning of *real*, where Austin argues that *real* is one of those words for which it is the negative uses that "wear the trousers," or, as we might say these days, "call the shots." To say that something is real leather or a real duck is not to say anything positive about leather or ducks but merely to exclude possible ways in which these things can fail to be genuine. Which possibilities are excluded will vary from context to context. Thus, Austin concludes that it is futile for philosophers to search for a property that is common to all things that are or could be called *real*. Hacking agrees. The word *real* is not ambiguous—its general function is the same in all cases—but its application varies from case to case. See Ian Hacking, *Representing and Intervening* (Cambridge: Cambridge University Press, 1983), 33.

Hacking includes an epistemological ingredient in his definition of entity realism because 'the whole issue would be idle' if we believe theoretical entities exist but do not believe that we could have warranted belief in their existence (ibid., 28).

Many realists are both entity realists and theory realists, and quite often entity realism is based on theory realism: if you believe that a theory is true, then you have good reasons for believing that the entities (e.g., objects, causal properties, structures) posited by the theory exist.² But Hacking holds that one can be an entity realist without being a theory realist, and that entity realism need not be based on theory realism. Hacking, by his own classification, is an entity realist: 'realism is not about theories and truth. The experimentalist need only be a realist about the entities used as tools.'³ Since other realists are both entity and theory realists, I will refer to Hacking's realism as theory-free entity realism.

Hacking's argument for theory-free entity realism rests on an analysis of scientific experimentation as well as some controversial metaphysical, epistemological, and semantic views. The experimental part of his argument is straightforward: experimentation largely proceeds without guidance from theories about the phenomena under study: 'much truly fundamental research precedes any relevant theory whatsoever' (*Representing and Intervening*, 158). Experimenters need to rely on some low-level generalizations about their instruments, as well as some theories about phenomena they are not studying, but they do not need to subscribe to full-blown scientific theories about the phenomena they are investigating in order to achieve their results. They need some theories to build instruments, but they do not need theories to use them. Hacking makes similar remarks about observation in science: observation, though guided by theory, need not be completely theory-laden: 'there have been important observations in the history of science, which have included no theoretical assumptions at all' (ibid., 176).

The metaphysical part of his argument turns on an anti-Humean analysis of causation and an instrumentalist account of laws and models. Causal claims, according to Humeans, are based on observed regularities in nature which we formulate as scientific laws. The causes we posit, according to Humeans, are not real; only regularities are real. Hacking reverses this doctrine: the causes are real but the regularities are not. Hacking follows Nancy Cartwright's account of causation and laws in physics and holds that theoretical laws of physics are best viewed as mere instruments ('Experimentation and Scientific Realism'). Theoretical laws are approximations; they are riddled with exceptions and based on idealizing assumptions. Moreover, physicists use a battery of models (some of them mutually inconsistent) to describe the same phenomena. Even though physical laws and models are best viewed as instruments, the entities described by the laws and models are not: the underlying causes are real; the regularities are not. Thus, physicists can make causal claims without

basing them on scientific laws or models. This view implies theory anti-realism, provided that one holds that laws and models are the backbone of scientific theories.

Hacking's metaphysics is supported, in part, by a causal theory of reference made popular by Hilary Putnam and Saul Kripke.* In *Representing and Intervening*, Hacking weds himself to Putnam's version of the causal theory of reference and meaning and in another paper he refers to 'Our debt to Hilary Putnam' (ibid., 157). Prior to the development of the causal theory of reference, many philosophers of science held what could be loosely described as a descriptivist theory of meaning.⁴ On the descriptivist view, words have a reference and a meaning (or a sense). Words or (phrases) refer to the objects they denote (if these objects exist) and the meanings of words are derived from descriptions we associate with them. Thus, 'water' means ' H_2O ' and we can use 'water' to denote (or pick out) objects, i.e. steam, ice, etc. But what happens to meaning when our descriptions change? If we once described water as simply 'clear liquid' and we now describe it as 'compound of two parts hydrogen and one part oxygen,' are we talking about the same thing? A whole host of semantic problems in science resulted from this approach to meaning, the most worrisome of these being Thomas Kuhn's infamous thesis that scientific theories are incommensurable.⁵ It is impossible to compare competing theories, according to Kuhn, because successive and competing theories are not even talking about the same things. Two theories, T1 and T2, might employ similar terms, such as 'mass' and 'velocity,' but the meanings of 'mass' and 'velocity' in T1 would not be the same as the meanings of 'mass' and 'velocity' in T2.†

These semantic problems are anathema for scientific realism because they lead to irrationalist conclusions, which, in turn, lead to anti-realist conclusions. If theories are incommensurable, then scientists cannot rationally adjudicate among theories. Even if we had theory-neutral methods and a theory-neutral evidence base, our language would be so theory-dependent that we could not objectively compare theories. We would 'choose' a new theory by learning its language and becoming committed to its assumptions, not by being swayed by the evidence. The anti-realist implication of this view is that reality is also theory-dependent: there is no

* For further discussion of the causal theory of reference and a criticism of Putnam's version of it, see D. H. Mellor, "Necessities and Universals in Natural Laws," in chapter 7 and the section "Mellor's Defense of the Regularity Theory" in the accompanying commentary.

† See Thomas S. Kuhn, "The Nature and Necessity of Scientific Revolutions," in chapter 2 and Ernest Nagel, "Issues in the Logic of Reductive Explanations" and Paul Feyerabend, "How to Be a Good Empiricist—A Plea for Tolerance in Matters Epistemological," in chapter 8.

objective reality, only reality within a given research community using a particular set of theories.⁶

Putnam's causal theory of reference saves philosophers of science from these (and other) difficulties, according to Hacking, by showing how scientists holding competing or successive theories 'may still be talking about the same thing' (*Representing and Intervening*, 75). The meaning of a word, according to Putnam, has four components: the word's syntactic marker, its semantic marker, its stereotype, and its referent.* The syntactic marker classifies the word grammatically, e.g. as a noun, verb, etc. . . . The semantic marker of a word indicates the general category of things it signifies, the stereotype is a description commonly associated with the word, and the referent of a word is that thing (or class of things) it denotes (or refers to) if it denotes at all. Stereotypes change—in science we frequently revise them—but referents, syntactic and semantic markers do not. Thus scientists who hold competing or successive theories can still be talking about the same things even though they describe those things in different ways.

This semantic view has important consequences for realism (in general) because it subverts worries about incommensurability: theories can be compared because different theories can refer to the same things. Scientists may use different descriptions (stereotypes) to talk about those things, but the things talked about exist independently of these descriptions. It also has important consequences for Hacking's version of realism because it shows how we might be anti-realists about theories but not about entities: theories, like stereotypes, come and go; but the entities they describe, i.e., their referents, remain the same. Hacking also hopes that this semantic view has important epistemological consequences, i.e., acceptance of theoretical entities is compatible with skepticism toward theories: 'one can believe in some entities without believing in any particular theory in which they are embedded' (*ibid.*, 29).

* Hilary Putnam, "The Meaning of 'Meaning'," in *Language, Mind, and Knowledge*, ed. K. Gunderson, vol. 7, *Minnesota Studies in the Philosophy of Science*, (Minneapolis: University of Minnesota Press, 1975); reprinted in Hilary Putnam, *Mind, Language, and Reality: Philosophical Papers*, vol. 2 (Cambridge: University of Cambridge Press, 1975), 215–71. See also chapter 6 of Ian Hacking's *Representing and Intervening*, for an exceptionally clear summary of Putnam's theory and a criticism of it. Hacking argues that, despite its many virtues for the scientific realist, Putnam's theory places so much emphasis on reference that it cannot do justice to cases (such as phlogiston and caloric) in which scientists agree about what a term means even though it has no referent, and other cases (such as the naming of muons and mesons) in which scientists reassign names on theoretical grounds.

■ | II

Now that we have a clearer understanding of the view Hacking attempts to defend, theory-free entity realism, we need to examine his argument for this position. Hacking's argument focuses on our ability to intervene in the world. Intervention is a kind of causal process in which cognizers use the world to *do, make, or change* things. Intervention (or intervening) contrasts with representation (or representing). Representation is a kind of causal process in which cognizers use language to represent the world. Representations (the products of the process of representation) are intended to be 'more or less public likenesses' of the world (ibid., 133). Representation is not action in the world; it is talk and thought about the world. Scientists represent the world through scientific theories and concepts.

Traditional arguments for (and against) realism have focused on representation, not intervention, according to Hacking. Their chief concern is to convince us that our theories and concepts do (or do not) accurately represent the world. Scientific realism is a very dubious doctrine if one focuses on representation, according to Hacking. Since we can never get outside of our representations and 'hook-up with the world,' the most plausible doctrine is 'some version of idealism' (ibid., 130). (Once Putnam realized that reference could not rescue realism from idealism, he took this option and became an internal realist, according to Hacking.)

Hacking seeks to reverse this trend by focusing on how we intervene in the world. Ordinary people intervene in the world by building bridges, chopping down trees, driving cars, and so forth. Scientists intervene in the world by experimenting. In experimentation, we take advantage of the world's causal structure in order to produce, control, and observe phenomena. If we use one part of the world to intervene in the world, then we are entitled to believe that our tool for intervention is real. (The structure that we use the tool on may still be regarded as only hypothetical.) Hacking's argument for entity realism now begins to fall in place. Since his writing is a bit enigmatic, highly suggestive, and sometimes vague, it will be useful to quote several passages in which he mounts this argument:

The best kinds of evidence for the reality of a postulated or inferred entity is that we can begin to measure or otherwise understand its causal powers. The best evidence, in turn, that we can have this kind of understanding is that we can set out, from scratch, to build machines that will work fairly reliably, taking advantage of this causal nexus. Hence, engineering, not theorizing, is the best proof of scientific realism about entities. ('Experimentation and Scientific Realism,' 170) [See original version in this chapter, page 1167]

Now, how does one alter the charge on the niobium ball? "Well, at that stage," said my friend, "we spray it with positrons to increase the charge or with electrons to decrease the charge." *So far as I'm concerned, if you can spray them then they are real.* (*Representing and Intervening*, 23)

I am told that nobody can yet manipulate a bunch of neutral bosons, if there are any. . . . When might they lose their hypothetical status and become commonplace reality like electrons? When we use them to investigate something else. (*ibid.*, 272)

In a nutshell, Hacking's argument for entity realism goes something like this:

- 1 We are entitled to believe that a theoretical entity is real if and only if we can use that entity to do things to the world.
- 2 We can use some theoretical entities, e.g., electrons, to do things to the world, e.g., change the charges on niobium balls.
- 3 Hence, we are entitled to believe that some theoretical entities, e.g., electrons, are real.

This argument, on the surface, is quite convincing. In experimentation we causally affect and manipulate things: we inject DNA into mouse embryos, we shoot neutrons at atoms, we drop balls off towers. It seems reasonable to believe that the objects we manipulate in experiments are real. When a theoretical entity becomes a tool—when it becomes something we can use and manipulate to achieve certain goals—it ceases to be merely hypothetical and becomes real. Thus, during Mendel's time genes were not real; after genetic engineering, they became real.

Hacking's argument for realism, though widely applicable, is not offered as an argument for an all-embracing realism. In other words, he holds that realism must be defended on a case-by-case basis. He defends what he calls 'realism in particular' not 'realism in general' (*ibid.*, 31). He would have us examine each science's experimental methods, tools, and its theoretical posits. If that science is able to use its posits as tools for experimentation, then the posits are real. If not, then they are merely hypothetical. An implication of this view is that one can be a realist about electrons but not about quarks or weak neutral currents. One can be a realist about DNA but not about species. Indeed, Hacking admits he is an anti-realist when it comes to the objects of astrophysics:

When we use entities as tools, as instruments of inquiry, we are entitled to regard them as real. But we cannot do that with objects of astrophysics. Astrophysics is almost the only human domain in which we have profound, intricate knowledge, and in which we can be no more than what van Fraassen calls constructive empiricists.⁷

Before criticizing Hacking's argument for realism, we should note that he does not dismiss other arguments for realism altogether: 'I did not say in my [*Representing and Intervening*] that an experimental argument is the only viable argument for scientific realism about unobserved entities. I said only that it is the most compelling one, and perhaps the *only* compelling one' (ibid., 560–1). Although Hacking hedges his bets in this and other passages, his disparaging remarks about arguments for realism which focus on representation instead of intervention indicate that he takes a dim view of other arguments for entity realism. Thus, we should read him as attempting to defend *the* argument for entity realism.

■ | III

Thus far, most of the critical discussion of Hacking's realism has focused on premise (1) in the argument sketched above. For instance, Dudley Shapere, in opposition to Hacking, argues for realism in astrophysics on the grounds that Hacking's sense of 'use' is ambiguous and vague.⁸ 'Use' could be interpreted actively as 'manipulate' or 'control' or more passively as 'employ' or 'exploit.' According to Shapere, only the active sense of 'use' supports anti-realism in astrophysics. But since (according to Shapere) Hacking has no argument for interpreting 'use' in this way, we can employ the passive sense of this slippery word to support realism in astrophysics. For instance, Shapere argues that since we can use gravitational lenses to measure distances, we are entitled to regard gravitational lenses as real even though we cannot do experiments on them (ibid.).

Although I agree with Shapere's critique of Hacking, it is not my aim in this essay to attack this aspect of Hacking's realism. Rather, I will discuss three other objections to his view. My first objection is that his argument is, despite his strong denials, another version of the success of science argument. This objection arises when we consider Hacking's second premise in the argument sketched above, viz., that we can use theoretical entities as tools. My concern is with how we support (or argue for) our claims to use theoretical entities as tools. These claims are far from obvious and need defending. Moreover, since they are causal claims, they demand fairly sophisticated arguments.

The first question to ask then is 'how do you support causal claims?' As David Hume pointed out long ago, we do not deduce causal claims from experience.⁹ Hume believed that our causal claims are based on induction: we observe regularities (or effects) in nature and infer causes from these regularities based on the inductive principle 'similar effects have similar causes.' Since Hume regarded induction as an unjustified form of reasoning, he held that our causal claims are unjustified as well. Although Hume's doubts about causality still influence all discussions of

causation, many philosophers now believe that we can justify causal claims through inductive or possibly abductive reasoning. Today, many scientists defend inductive arguments for causal connections based on statistical reasoning.¹⁰ Many philosophers also hold that one can base causal claims on inference to the best explanation, or abductive reasoning: A causes B if and only if the hypothesis 'A causes B' provides the best explanation of the evidence.

Hacking, as we saw earlier, does not believe that we can base causal claims on observed regularities. Thus, he would not (or should not!) take an inductivist approach to the justification of causal claims. But if he dismisses this approach, the only viable alternative is an abductivist approach. Hacking does not commit himself to any particular methods of scientific reasoning, but his argument for experimental realism bears a strong resemblance to other abductivist arguments for realism, such as the infamous success of science argument. According to J. J. C. Smart's version of this argument, realism is the best explanation of the predictive and explanatory success of particular scientific theories.¹¹ Smart argues it would be an incredible coincidence—a 'cosmic accident'—if a false theory made accurate predictions, gave good explanations, and so forth. Hence, we should believe that theories which exhibit explanatory and predictive success are true. According to Richard Boyd's version, realism is the best explanation of the success (or instrumental reliability) of scientific methodology.¹² Boyd claims that it would also be an incredible coincidence if successful sciences, which deploy theory-dependent methods, were not making progressive approximation toward the truth. Thus, we should believe that the theories of successful sciences are approximately true and genuinely refer. In both arguments success is an acknowledged fact and realism is supposed to be the best explanation of this fact.

Hacking's argument is also a 'success of science argument' for realism. While Smart and Boyd focus on the success of theoretical science, viz., predictive and explanatory success, Hacking focuses on the success of experimental science. Success in experimentation is equated with the ability to reliably manipulate or control our instruments in order to produce desired effects ('Experimentation and Scientific Realism'). A successful experimenter can also dampen 'noise' and undesired effects. To reiterate an earlier quotation, Hacking says that the best evidence we can have that we use an entity as a tool is that we can 'build machines that will work fairly reliably' (ibid., 170 [1167]). Although Hacking denies that experimenters need to accept any theories about the phenomena they study, he acknowledges that they must rely on some background assumptions and low-level generalizations (*Representing and Intervening*, 149–67). For instance, an experimenter who used electrons to do things, would at least have to assume that electrons are a part of the world's causal nexus. Hacking's argument for premise (2) above now takes shape:

- A Our scientific instruments work fairly reliably in producing desired effects, e.g., changing the charge on niobium balls.
- B The best explanation of this reliability is that these instruments take advantage of a (real) causal nexus in the world. That is, that we are using the instruments to do things in the world.
- C According to our low-level generalizations (theories?) and other assumptions, our instruments take advantage of some theoretical entities in producing their effects.
- D Hence, we can use some theoretical entities, e.g., electrons, to do things in the world, e.g., change the charge on niobium balls.

Hacking's argument, like other success of science arguments, also depends on our recognizing an incredible coincidence that cries out for explanation. The coincidence for the experimenter is that the instruments work reliably in producing desired effects. It would be an incredible coincidence if these instruments worked reliably but were not taking advantage of the world's causal nexus. It would be an incredible coincidence if we managed to use an entity as a tool for inquiry that did not, in fact, exist. Thus, we should believe that the entities we are using exist and that we are taking advantage of the world's causal nexus in order to avoid accepting an incredible coincidence.

Hacking repeatedly asserts that his argument for experimental realism is not just another version of the infamous success of science argument:

We "infer the best explanation" that the theory is true. The common cause of the phenomena must be the theoretical entities postulated by the theory. As an argument for scientific realism this idea has produced much debate. So it must seem as if my talk of coincidence puts me in the middle of an ongoing feud. Not so! My argument is much more localized. (ibid., 202)

Once upon a time the best reason for thinking that there are electrons might have been success in explanation. . . . Luckily we no longer have to pretend to infer from explanatory success. (ibid., 271–2)

In these passages Hacking asserts that his argument for realism is not a success of science argument, but he offers no sustained argument to support this assertion. The closest thing I can find to an argument in his work is his claim that experimenters use entities as tools; they do not use them to explain: 'Prescott *et al.* don't explain phenomena with electrons. They know how to use them' (ibid., 272).

The problem with this line of defense is that one cannot *rationaly* claim to use a theoretical entity as a tool of inquiry without some evidence, argument, or justification. Our claims about using or manipulating theoretical entities should not be dogmatic assertions: they require some jus-

tifications. These justifications can employ either inductive or abductive arguments. If our claims to use theoretical entities as tools are not supported by inductive arguments, as Hacking appears to maintain, then they must be supported by abductive ones. Hence, if one regards an entity as a tool of inquiry, one must also claim that its place in the world's causal structure explains some phenomena.

Putting my reply in a slightly different light, we can distinguish between two additional senses of the word 'use': an epistemic sense and a non-epistemic one. In the epistemic senses of 'use,' one can use an entity only if one claims to know something about it; in the non-epistemic sense of 'use,' one can use an entity without claiming to know anything about it. When we say that we 'use an equation to calculate probabilities' we are using 'use' in the epistemic sense; when we say that we 'use oxygen in respiration,' we are using 'use' in the non-epistemic sense. My argument is that experimentation is, in part, an epistemic activity, and that experimental uses are, in part, epistemic uses. Some activities, such as breathing, can be viewed as non-epistemic: I can breath (or use) oxygen without claiming to know anything about it. We can use many things in this world while remaining completely ignorant about them and having no evidence which might have some bearing on their causal properties. But experimentation is not like this. One cannot use a theoretical entity in an experiment without claiming to know something about that entity and claiming to have some justification for believing that it has specific causal properties.

Thus, using an entity as a tool in experiment also commits us to regarding it as an entity we can use to explain and predict phenomena. But an entity does not explain anything by itself; it explains phenomena within the context of a theory that describes it. Hence, if we believe that we can use a theoretical entity to explain and predict phenomena, we must also believe that our theories which describe the entity are at least approximately true, since theories that are not at least approximately true cannot explain. (More will be made of this point in the next section.)

One reason why Hacking is against inference to the best explanation is that he agrees with the charge made by Bas van Fraassen and others that this form of argument is unscientific because explanation is a hopelessly vague or interest-relative notion.¹³ However, some recent theories of explanation attempt to counter this charge by articulating causal accounts of explanation.¹⁴ In an analysis of inference to the best explanation, Peter Lipton proposes an approach to abduction that employs a causal model of explanation, and thus may make this form of inference appear to be less vague or interest-relative.¹⁵ If these approaches to explanation and abduction prove to be fruitful, they may offer plausible arguments for realism, and Hacking's experimental realism can be viewed as a contribution to this trend.

■ | IV

My second and third criticisms of Hacking's view are that he fails to make a plausible case for theory-free, entity realism. His view is on shaky philosophical ground because it does not allow experimental realists to have knowledge about the entities they investigate. Furthermore, it does not accurately portray experimental practice.

Hacking's theory-free, entity realism is a metaphysically consistent position—one *can* believe in theoretical entities without believing in the theories in which they are embedded—but it is not a reasonable position. It is not a reasonable position because it gives the experimenter belief (and perhaps even true belief) in theoretical entities without justified belief. Given that justification is a necessary condition for knowledge, theory-free entity realism cannot yield knowledge about theoretical entities. And if Hacking's realism is a position which does not yield knowledge, then it would be, as he says, 'idle.'

In order to explore this criticism, we need to return to Putnam's semantic theory. This theory, as we saw earlier, provides the semantic backbone for Hacking's metaphysics and epistemology. It enables Hacking to claim that we are justified in believing in a theoretical entity without believing in the theory in which it is embedded because we can continue to successfully refer to that entity even when our theories about that entity undergo radical changes. While this view establishes the cogency of Hacking's metaphysical part of theory-free entity realism, it does not go very far in supporting the epistemological aspect of this view. Theory-free entity realism is on epistemologically solid ground—it can give us knowledge—only if we know that the entities to which we refer are natural kinds; that is, we know that they are fundamental parts of the world's causal structure. However, if we do not know that a theoretical entity is a natural kind, then we cannot claim to continue to successfully refer to that entity when our theories about it change. If a theoretical entity is not a natural kind, then it is an artifact of our theories and classificatory systems. If an entity is a mere artifact, then we cannot have knowledge about that entity.

Some examples will serve to clarify this point. Phlogiston was a theoretical entity used to explain combustion, rusting, metabolism, and other chemical phenomena (see Kuhn). For instance, according to phlogiston theory, during combustion, the burning material emitted smoke and phlogiston—the 'fire substance'—and left behind ashes. This view was supported, in part, by the fact that materials lose weight during combustion, and part of this loss was attributed to the loss of phlogiston. The theory also predicted that materials would lose phlogiston (and thus weight) during other chemical processes. However, experiments showed that materials did not lose weight as the theory had predicted but actually gained weight. Since defenders of the phlogiston theory held that mass is conserved in

all chemical reactions, they proposed that phlogiston could have no mass or even negative mass. But these ad hoc attempts to save the theory failed, and phlogiston was no longer regarded as a real entity. Phlogiston was not a natural kind; it was merely an artifact of specific chemical theories. Theorists eventually explained the phenomena explained by phlogiston by positing other theoretical entities, such as oxygen and carbon dioxide.

Gregor Mendel posited the existence of hereditary factors (or Mendelian genes) to explain hereditary phenomena. Mendel formulated three laws that governed these entities. The laws were used to explain the phenotypic similarities and differences between parents and offspring and to predict hereditary phenomena. During this century, genetics discovered that Mendel's laws are fraught with exceptions and that the functions performed by Mendelian genes are actually performed by several different kinds of entities, such as cistrons, mutons, and recons, and that the functions performed by these entities are actually performed by DNA.¹⁶ Although geneticists still use the term 'gene,' Mendelian genes can no longer be viewed as distinct, natural kinds. Unlike phlogiston, Mendelian genes are not fictional entities, but they are artifacts of Mendelian theory, nonetheless. Mendelian genes have been, in a sense, split into many different parts by modern genetic theories.

These two examples are neither trivial nor obscure; the history of science contains many episodes where theoretical entities have been shown to be mere artifacts.¹⁷ The question we need to put before Hacking is this: how do we *know* that theoretical entities that we talk about actually exist? How do we know that electrons will not one day be shown to be artifacts of current physical theories? For all we know, electrons could be like Mendelian genes: their effects might really be produced by other entities. Or even worse, they might turn out to be like phlogiston and not exist in any way at all.

Hacking has a simple reply to these questions: we know that electrons (and other theoretical entities) exist because we can refer to them. Our theories about electrons may change, but we can still refer to the same things. But this reply simply begs the question, since we cannot claim to successfully refer to a theoretical entity unless we are already justified in claiming that the entity is a natural kind. Reference, as Hacking himself admits, is not a magical 'sky-hook' which can save us from anti-realism (Hacking, *Representing and Intervening*, 130).

The next issue we need to address, then, is how we can know that a theoretical entity is a natural kind. For if we can know that a theoretical entity is a natural kind, then we can know that we can successfully refer to that entity and that we have knowledge about that entity. Anti-realists, of course, might assert that we can never know that a theoretical entity is a natural kind and that we thus cannot have knowledge about theoretical entities. But some realists, such as Boyd, propose that we can know that an entity is a natural kind provided that we have a well-confirmed, ex-

planatorily successful, approximately true, theory that posits the existence of that entity. We can know that electrons are natural kinds because they are posits of our current, highly successful, theories. This suggestion also leads us back to a conclusion reached in the last section, that the experimental realist is committed to the approximate truth of the theories in which her entities are embedded. Ironically, this realist alternative also brings Hacking back into the traditional representationalist approach to realism.

The trouble with the alternative, of course, is this is precisely that view that Hacking disdains. He wants to maintain that we have knowledge about theoretical entities without believing in any theories in which they are embedded and he wants to dodge the success of science argument for realism. He might try to wriggle his way out of this pigeonhole by claiming that we can know that a theoretical entity is a natural kind if we can use it as a tool for inquiry. But this move does not free him from theoretical bindings. As we saw earlier, we cannot claim to use a theoretical entity as a tool for inquiry unless we also claim that it explains our experimental successes. But how do we explain those successes? By appealing to *theories* that describe the causal processes and structures used in our experiments. Thus, experimenters are justified in believing in electrons because electrons are posited by theories that explain experimental successes.

Hacking, like Houdini, might have one other way to escape this theoretical cage. He could maintain that experimenters need to rely on low-level generalizations about the entities they study, but they do not need to commit themselves to full-blown scientific theories about these entities. Although Hacking does not say what the difference might be between a low-level generalization and a full-blown theory, one commonly accepted way of distinguishing between theories and low-level generalizations is to claim that theories must contain genuine, scientific laws.¹⁸ Scientific laws are unrestricted, universal, generalizations that possess 'nomic necessity' (ibid.).*

There are two problems with this reply. First, it is based on an overly restrictive view of theories. This view is overly restrictive because it implies that the only 'genuine' theories will be found in physics and chemistry. But this position does not accurately reflect scientific practice and belittles theories in the biological, earth, and human sciences. It makes perfect sense to talk about theories of fetal development, plate tectonics, stellar evolution, and animal aggression. A theory need not contain any unrestricted, universal generalizations; there can be theories that apply only to particular things or systems, or which contain only low-level generalizations.

Second, even if we accept this overly restrictive view of theories, it turns out that experimenters actually use full-blown scientific theories about the very entities they study. While one could, for the sake of phil-

* For more on laws and nomic necessity, see chapter 7.

osophical argument, claim that experimenters only need rough generalizations, this philosophical fantasy has no basis in experimental practice. Experimenters do not operate without genuine scientific theories and laws about the phenomena they investigate: the gulf between experiment and theory is not nearly as large as Hacking supposes. Indeed, during the scientific revolution there was no sharp distinction between theorists and experimenters: many of the best theorists, such as Galileo, Newton, Harvey, and Boyle, were also excellent experimenters. As scientists have become more specialized, a gap has developed between theorists and experimenters. However, even today these two different communities share some common reference points. A person running experiments with a particle accelerator may not be aware of the latest developments in theoretical physics, but he (or she) is likely to be familiar with most of the commonly accepted background theories in physics, including some theories about the particles he (or she) is studying. Experimenters and theorists both obtain graduate degrees in their chosen fields and are thus both familiar with a common core of background theories. They also continue to communicate about scientific ideas and instruments as they pursue their different career paths. (Many of these claims about experimentation are empirical points which require further substantiation. Since I do not have the space to provide that support here, I refer the reader to the work of Allan Franklin.¹⁹)

The upshot of this discussion is that Hacking's theory-free entity realism can be attacked on both philosophical and practical grounds. The philosophical problem with this position is that it cannot give us knowledge about theoretical entities; the practical problem with it is that it does not accurately reflect scientific practice.

■ | V

Hacking is right to point out that philosophers of science have placed too much emphasis on scientific representation and have for too long neglected scientific intervention. Given the neglect of experimentation and technology, Hacking's realism is a refreshing and insightful approach. However, Hacking is wrong to think that questions about representation are irrelevant to arguments for realism; the three critiques developed in this essay illustrate this point. Although Hacking eschews theory realism and inference to the best explanation, his experimental realism does not elude either of these traditional realist concerns. My criticism, then, might be properly regarded as a reinterpretation of Hacking's experimental realism, since I do not attack his view *per se*. I think that experimental realism is a plausible and interesting position, but it is not as non-traditional, non-theoretical, or non-abductive as Hacking assumes it is.²⁰

■ | Notes

1. Ian Hacking, *Representing and Intervening* (Cambridge: Cambridge University Press 1983).
2. Wilfrid Sellars, *Science, Perception and Reality* (New York: Humanities Press 1963). According to Sellars, 'to have good reason for holding a theory is *ipso facto* to have good reason for holding that the entities postulated by the theory exist' (97).
3. Hacking, 'Experimentation and Scientific Realism,' in J. Leplin, ed., *Scientific Realism* (Berkeley and Los Angeles: University of California Press 1984), 154 [p. 1153 this volume]. Hacking aligns himself with Nancy Cartwright, who is also an entity realist. She holds that the theoretical laws of physical theories are false, but that we can have good reasons for believing that entities posited by those theories exist. She also believes, like Hacking, that we can have good reasons for believing that theoretical entities exist when we can causally interact with them. See her book *How the Laws of Physics Lie* (Oxford: Oxford University Press 1983).
It should also be noted that by 'theoretical entity' Hacking means nothing more than an entity that cannot be observed with the unaided senses, and that a theory, according to Hacking, is not a mere model of the phenomena or a collection of low-level generalizations or phenomenological laws; a theory contains some high-level generalizations or fundamental laws. Thus Hacking's account of theories also bears a strong resemblance to Cartwright's view.
4. Bertrand Russell, *Introduction to Mathematical Philosophy* (London: George Allen and Unwin 1919), 167–80.
5. Thomas Kuhn, *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press 1962).
6. W. H. Newton-Smith, *The Rationality of Science* (London: Routledge and Kegan Paul 1981).
7. Hacking, 'Extragalactic Reality: The Case of Gravitational Lensing,' *Philosophy of Science* 56 (1989), 578.
8. Dudley Shapere, 'Astronomy and Anti-Realism,' *Philosophy of Science* 60 (1993) 134–50.
9. David Hume, *An Inquiry Concerning Human Understanding* (Indianapolis: Bobbs-Merrill 1955).
10. Richard Johnson and Gouri Bhattacharya, *Statistics: Principles and Methods* (New York: John Wiley and Sons 1985).
11. J. J. C. Smart, *Between Science and Philosophy* (New York: Random House 1968).
12. Richard Boyd, 'The Current Status of Scientific Realism,' in J. Leplin, ed., *Scientific Realism*.
13. Bas van Fraassen, *The Scientific Image* (Oxford: Clarendon Press 1980).
14. Wesley Salmon, *Scientific Explanation and the Causal Structure of the World* (Princeton: Princeton University Press 1984).

15. Peter Lipton, *Inference to the Best Explanation* (London: Routledge 1991).
16. Alexander Rosenberg, *The Structure of Biological Science* (Cambridge: Cambridge University Press 1985).
17. Larry Laudan, 'A Confutation of Convergent Realism,' *Philosophy of Science* 48 (1981) 19–49. [Excerpted in this chapter.]
18. Carl Hempel, *Philosophy of Natural Science* (Englewood Cliffs, NJ: Prentice-Hall 1966).
19. Allan Franklin, *The Neglect of Experiment* (Cambridge: Cambridge University Press 1986).
20. I am grateful to Allan Franklin and an anonymous reviewer from the *Canadian Journal of Philosophy* for helpful suggestions and comments.

The Natural Ontological Attitude

Let us fix our attention out of ourselves as much as possible; let us chace our imagination to the heavens, or to the utmost limits of the universe; we never really advance a step beyond ourselves, nor can conceive any kind of existence, but those perceptions, which have appear'd in that narrow compass. This is the universe of the imagination, nor have we any idea but what is there produced.

—Hume, *Treatise*, Book 1, Part II, Section VI

Realism is dead. Its death was announced by the neopositivists who realized that they could accept all the results of science, including all the members of the scientific zoo, and still declare that the questions raised by the existence claims of realism were mere pseudoquestions. Its death was hastened by the debates over the interpretation of quantum theory, where Bohr's nonrealist philosophy was seen to win out over Einstein's passionate realism. Its death was certified, finally, as the last two generations of physical scientists turned their backs on realism and have managed, nevertheless, to do science successfully without it. To be sure, some recent philosophical literature . . . has appeared to pump up the ghostly shell and to give it new life. But I think these efforts will eventually be seen and understood as the first stage in the process of mourning, the stage of denial. . . . But I think we shall pass through this first stage and into that of acceptance, for realism is well and truly dead, and we have work to get on with, in identifying a suitable successor. To aid that work I want to do three things in this essay. First, I want to show that the arguments in favor of realism are not sound, and that they provide no rational support for belief in realism. Then, I want to recount the essential role of nonrealist attitudes for the development of science in this century, and thereby (I hope) to loosen the grip of the idea that only realism provides a progressive philosophy of science. Finally, I want to sketch out what seems to me a viable nonrealist position, one that is slowly gathering support and that seems a decent philosophy for postrealist times.¹

■ | Arguments for Realism

Recent philosophical argument in support of realism tries to move from the success of the scientific enterprise to the necessity for a realist account of its practice. As I see it, the arguments here fall on two distinct levels. On the ground level, as it were, one attends to particular successes; such as novel, confirmed predictions, striking unifications of disparate-seeming phenomena (or fields), successful piggybacking from one theoretical model to another, and the like. Then, we are challenged to account for such success, and told that the best and, it is slyly suggested, perhaps, the *only* way of doing so is on a realist basis. I do not find the details of these ground-level arguments at all convincing. Larry Laudan has provided a forceful and detailed analysis which shows that not even with a lot of hand waving (to shield the gaps in the argument) and charity (to excuse them) can realism itself be used to explain the very successes to which it invites our attention.² But there is a second level of realist argument, the methodological level, that derives from Popper's attack on instrumentalism as inadequate to account for the details of his own, falsificationist methodology. Arguments on this methodological level have been skillfully developed by Richard Boyd,³ and by one of the earlier Hilary Putnams.⁴ These arguments focus on the methods embedded in scientific practice, methods teased out in ways that seem to me accurate and perceptive about ongoing science. We are then challenged to account for why these methods lead to scientific success and told that the best, and (again) perhaps, the only truly adequate way of explaining the matter is on the basis of realism.

I want to examine some of these methodological arguments in detail to display the flaws that seem to be inherent in them. But first I want to point out a deep and, I think, insurmountable problem with this entire strategy of defending realism, as I have laid it out above. To set up the problem, let me review the debates in the early part of this century over the foundations of mathematics, the debates that followed Cantor's introduction of set theory. There were two central worries here, one over the meaningfulness of Cantor's hierarchy of sets insofar as it outstripped the number-theoretic content required by Kronecker (and others); the second worry, certainly deriving in good part from the first, was for the consistency (or not) of the whole business. In this context, Hilbert devised a quite brilliant program to try to show the consistency of a mathematical theory by using only the most stringent and secure means. In particular, if one were concerned over the consistency of set theory, then clearly a set-theoretic proof of consistency would be of no avail. For if set theory were inconsistent, then such a consistency proof would be both possible and of no significance. Thus, Hilbert suggested that finite constructivist means, satisfactory even to Kronecker (or Brouwer) ought to be employed in meta-

mathematics. Of course, Hilbert's program was brought to an end in 1931, when Gödel showed the impossibility of such a stringent consistency proof. But Hilbert's idea was, I think, correct even though it proved to be unworkable. Metatheoretic arguments must satisfy more stringent requirements than those placed on the arguments used by the theory in question, for otherwise the significance of reasoning about the theory is simply moot. I think this maxim applies with particular force to the discussion of realism.

Those suspicious of realism, from Osiander* to Poincaré and Duhem to the 'constructive empiricism' of van Fraassen,⁵ have been worried about the significance of the explanatory apparatus in scientific investigations. While they appreciate the systematization and coherence brought about by scientific explanation, they question whether acceptable explanations need to be true and, hence, whether the entities mentioned in explanatory principles need to exist.⁶ Suppose they are right. Suppose, that is, that the usual explanation-inferring devices in scientific practice do not lead to principles that are reliably true (or nearly so), nor to entities whose existence (or near-existence) is reliable. In that case, the usual abductive methods that lead us to good explanations (even to 'the best explanation') cannot be counted on to yield results even approximately true. But the strategy that leads to realism, as I have indicated, is just such an ordinary sort of abductive inference. Hence, if the nonrealist were correct in his doubts, then such an inference to realism as the best explanation (or the like), while possible, would be of no significance—exactly as in the case of a consistency proof using the methods of an inconsistent system. It seems, then, that Hilbert's maxim applies to the debate over realism: to argue for realism one must employ methods more stringent than those in ordinary scientific practice. In particular, one must not beg the question as to the significance of explanatory hypotheses by assuming that they carry truth as well as explanatory efficacy.

There is a second way of seeing the same result. Notice that the issue over realism is precisely the issue as to whether we should believe in the reality of those individuals, properties, relations, processes, and so forth, used in well-supported explanatory hypotheses. Now what is the hypothesis of realism, as it arises as an explanation of scientific practice? It is just the hypothesis that our accepted scientific theories are approximately true, where "being approximately true" is taken to denote an extratheoretical

* Andreas Osiander was a Lutheran theologian who contributed a short, anonymous preface to the first edition of Copernicus's *De Revolutionibus* (1543), arguing that Copernicus's theory should be regarded merely as a means of "saving the appearances," not as a realistic hypothesis. Copernicus, himself a realist, died before he could see the final version of his book, and Osiander supervised its publication. Osiander's imposture was detected and publicized by Kepler, but few astronomers had been deceived by it.

relation between theories and the world. Thus, to address doubts over the reality of relations posited by explanatory hypotheses, the realist proceeds to introduce a further explanatory hypothesis (realism), itself positing such a relation (approximate truth). Surely anyone serious about the issue of realism, and with an open mind about it, would have to behave inconsistently if he were to accept the realist move as satisfactory.

Thus, both at the ground level and at the level of methodology, no support accrues to realism by showing that realism is a good hypothesis for explaining scientific practice. If we are open-minded about realism to begin with, then such a demonstration (even if successful) merely begs the question that we have left open ("need we take good explanatory hypotheses as true?"). Thus, Hilbert's maxim applies, and we must employ patterns of argument more stringent than the usual abductive ones. What might they be? Well, the obvious candidates are patterns of induction leading to empirical generalizations. But, to frame empirical generalizations, we must first have some observable connections between observables. For realism, this must connect theories with the world by way of approximate truth. But no such connections are observable and, hence, suitable as the basis for an inductive inference. I do not want to labor the points at issue here. They amount to the well-known idea that realism commits one to an unverifiable correspondence with the world. So far as I am aware, no recent defender of realism has tried to make a case based on a Hilbert strategy of using suitably stringent grounds and, given the problems over correspondence, it is probably just as well.

The strategy of arguments to realism as a good explanatory hypothesis, then, *cannot* (logically speaking) be effective for an open-minded nonbeliever. But what of the believer? Might he not, at least, show a kind of internal coherence about realism as an overriding philosophy of science, and should that not be of some solace, at least for the realist?⁷ Recall, however, the analogue with consistency proofs for inconsistent systems. That sort of harmony should be of no solace to anyone. But for realism, I fear, the verdict is even harsher. For, so far as I can see, the arguments in question just do not work, and the reason for that has to do with the same question-begging procedures that I have already identified. Let me look closely at some methodological arguments in order to display the problems.

A typical realist argument on the methodological level deals with what I shall call the problem of the "small handful." It goes like this. At any time, in a given scientific area, only a small handful of alternative theories (or hypotheses) are in the field. Only such a small handful are seriously considered as competitors, or as possible successors to some theory requiring revision. Moreover, in general, this handful displays a sort of family resemblance in that none of these live options will be too far from the previously accepted theories in the field, each preserving the well-confirmed features of the earlier theories and deviating only in those as-

pects less confirmed. Why? Why does this narrowing down of our choices to such a small handful of cousins of our previously accepted theories work to produce good successor theories?

The realist answers this as follows. Suppose that the already existing theories are themselves approximately true descriptions of the domain under consideration. Then surely it is reasonable to restrict one's search for successor theories to those whose ontologies and laws resemble what we already have, especially where what we already have is well confirmed. And if these earlier theories were approximately true, then so will be such conservative successors. Hence, such successors will be good predictive instruments; that is, they will be successful in their own right.

The small-handful problem raises three distinct questions: (1) why only a small handful out of the (theoretically) infinite number of possibilities? (2) why the conservative family resemblance between members of the handful? and (3) why does the strategy of narrowing the choices in this way work so well? The realist response does not seem to address the first issue at all, for even if we restrict ourselves just to successor theories resembling their progenitors, as suggested, there would still, theoretically, always be more than a small handful of these. To answer the second question, as to why conserve the well-confirmed features of ontology and laws, the realist must suppose that such confirmation is a mark of an approximately correct ontology and approximately true laws. But how could the realist possibly justify such an assumption? Surely, there is no valid inference of the form "*T* is well confirmed; therefore, there exist objects pretty much of the sort required by *T* and satisfying laws approximating to those of *T*." Any of the dramatic shifts of ontology in science will show the invalidity of this schema. For example, the loss of the ether from the turn-of-the-century electrodynamic theories demonstrates this at the level of ontology, and the dynamics of the Rutherford-Bohr atom vis-à-vis the classical energy principles for rotating systems demonstrates it at the level of laws. Of course, the realist might respond that there is no question of a strict inference between being well confirmed and being approximately true (in the relevant respects), but there is a probable inference of some sort. But of what sort? Certainly there is no probability relation that rests on inductive evidence here. For there is no independent evidence for the relation of approximate truth itself; at least, the realist has yet to produce any evidence that is independent of the argument under examination. But if the probabilities are not grounded inductively, then how else? Here, I think the realist may well try to fall back on his original strategy, and suggest that being approximately true provides the best explanation for being well confirmed. This move throws us back to the ground-level realist argument, the argument from specific success to an approximately true description of reality, which Laudan has criticized. I should point out, before looking at the third question, that if this last move is the one the realist wants to make, then his success at the methodological

level can be no better than his success at the ground level. If he fails there, he fails across the board.

The third question, and the one I think the realist puts most weight on, is why does the small-handful strategy work so well. The instrumentalist, for example, is thought to have no answer here. He must just note that it does work well, and be content with that. The realist, however, can explain why it works by citing the transfer of approximate truth from predecessor theories to the successor theories. But what does this explain? At best, it explains why the successor theories cover the same ground as well as their predecessors, for the conservative strategy under consideration assures that. But note that here the instrumentalist can offer the same account: if we insist on preserving the well-confirmed components of earlier theories in later theories, then, of course the later ones will do well over the well-confirmed ground. The difficulty, however, is not here at all but rather is in how to account for the successes of the later theories in new ground or with respect to novel predictions, or in overcoming the anomalies of the earlier theories. And what can the realist possibly say in this area except that the theorist, in proposing a new theory, has happened to make a good guess? For nothing in the approximate truth of the old theory can guarantee (or even make it likely) that modifying the theory in its less-confirmed parts will produce a progressive shift. The history of science shows well enough how such tinkering succeeds only now and again, and fails for the most part. This history of failures can scarcely be adduced to explain the occasional success. The idea that by extending what is approximately true one is likely to bring new approximate truth is a chimera. It finds support neither in the logic of approximate truth nor in the history of science. The problem for the realist is how to explain the *occasional success* of a strategy that *usually fails*.⁸ I think he has no special resources with which to do this. In particular, his usual fallback onto approximate truth provides nothing more than a gentle pillow. He may rest on it comfortably, but it does not really help to move his cause forward.

The problem of the small handful raises three challenges: why small, why narrowly related, and why does it work? The realist has no answer for the first of these, begs the question as to the truth of explanatory hypotheses on the second, and has no resources for addressing the third. For comparison, it may be useful to see how well his archenemy, the instrumentalist, fares on the same turf. The instrumentalist, I think, has a substantial basis for addressing the questions of smallness and narrowness, for he can point out that it is extremely difficult to come up with alternative theories that satisfy the many empirical constraints posed by the instrumental success of theories already in the field. Often it is hard enough to come up with even one such alternative. Moreover, the common apprenticeship of scientists working in the same area certainly has the effect of narrowing down the range of options by channeling thought into the commonly accepted categories. If we add to this the instrumentally justified rule, "If

it has worked well in the past, try it again," then we get a rather good account, I think, of why there is usually only a small and narrow handful. As to why this strategy works to produce instrumentally successful science, we have already noted that for the most part it does not. Most of what this strategy produces are failures. It is a quirk of scientific memory that this fact gets obscured, much as do the memories of bad times during a holiday vacation when we recount all our "wonderful" vacation adventures to a friend. Those instrumentalists who incline to a general account of knowledge as a social construction can go further at this juncture, and lean on the sociology of science to explain how the scientific community "creates" its knowledge. I am content just to back off here and note that over the problem of the small handful, the instrumentalist scores at least two out of three, whereas the realist, left to his own devices, has struck out.⁹

I think the source of the realist's failure here is endemic to the methodological level, infecting all of his arguments in this domain. It resides, in the first instance, in his repeating the question-begging move from explanatory efficacy to the truth of the explanatory hypothesis. And in the second instance, it resides in his twofold mishandling of the concept of approximate truth: first, in his trying to project from some body of assumed approximate truths to some further and novel such truths, and second, in his needing genuine access to the relation of correspondence. There are no general connections of this first sort, however, sanctioned by the logic of approximate truth, nor secondly, any such warranted access. However, the realist must pretend that there are, in order to claim explanatory power for his realism. We have seen those two agents infecting the realist way with the problem of the small handful. Let me show them at work in another methodological favorite of the realist, the "problem of conjunctions."

The problem of conjunctions is this. If T and T' are independently well-confirmed, explanatory theories, and if no shared term is ambiguous between the two, then we expect the conjunction of T and T' to be a reliable predictive instrument (provided, of course, that the theories are not mutually inconsistent). Why? challenges the realist, and he answers as follows. If we make the realist assumption that T and T' , being well confirmed, are approximately true of the entities (etc.) to which they refer, and if the unambiguity requirement is taken realistically as requiring a domain of common reference, then the conjunction of the two theories will also be approximately true and, hence, it will produce reliable observational predictions. Q.E.D.

But notice our agents at work. First, the realist makes the question-begging move from explanations to their approximate truth, and then he mistreats approximate truth. For nothing in the logic of approximate truth sanctions the inference from " T is approximately true" and " T' is approximately true" to the conclusion that the conjunction " $T \cdot T'$ " is approxi-

mately true. Rather, in general, the tightness of an approximation dissipates as we pile on further approximations. If T is within ϵ , in its estimation of some parameter, and T' is also within ϵ , then the only general thing we can say is that the conjunction will be within 2ϵ of the parameter. Thus, the logic of approximate truth should lead us to the opposite conclusion here; that is, that the conjunction of two theories is, in general, *less* reliable than either (over their common domain). But this is neither what we expect nor what we find. Thus, it seems quite implausible that our actual expectations about the reliability of conjunctions rest on the realist's stock of approximate truths.

Of course, the realist could try to retrench here and pose an additional requirement of some sort of uniformity on the character of the approximations, as between T and T' .¹⁰ It is difficult to see how the realist could do this successfully without making reference to the distance between the approximations and "the truth." For what kind of internalist requirement could possibly insure the narrowing of this distance? But the realist is in no position to impose such requirements, since neither he nor anyone else has the requisite access to "the truth." Thus, whatever uniformity-of-approximation condition the realist might impose, we could still demand to be shown that this leads closer to the truth, not farther away. The realist will have no demonstration, except to point out to us that it all works (sometimes!). But that was the original puzzle.¹¹ Actually, I think the puzzle is not very difficult. For surely, if we do not entangle ourselves with issues over approximation, there is no deep mystery as to why two compatible and successful theories lead us to expect their conjunction to be successful. For in forming the conjunction, we just add the reliable predictions of one onto the reliable predictions of the other, having antecedently ruled out the possibility of conflict.

There is more to be said about this topic. In particular, we need to address the question as to why we expect the logical gears of the two theories to mesh. However, I think that a discussion of the realist position here would only bring up the same methodological and logical problems that we have already uncovered at the center of the realist argument.

Indeed, this schema of knots in the realist argument applies across the board and vitiates every single argument at the methodological level. Thus my conclusion here is harsh, indeed. The methodological arguments for realism fail, even though, were they successful, they would still not support the case. For the general strategy they are supposed to implement is just not stringent enough to provide rational support for realism. In the next two sections, I will try to show that this situation is just as well, for realism has not always been a progressive factor in the development of science and, anyway, there is a position other than realism that is more attractive.

■ | Realism and Progress

If we examine the two twentieth-century giants among physical theories, relativity and the quantum theory, we find a living refutation of the realist's claim that only his view of science explains its progress, and we find some curious twists and contrasts over realism as well. The theories of relativity are almost singlehandedly the work of Albert Einstein. Einstein's early positivism and his methodological debt to Mach (and Hume) leap right out of the pages of the 1905 paper on special relativity.¹² The same positivist strain is evident in the 1916 general relativity paper as well, where Einstein (in Section 3 of that paper) tries to justify his requirement of general covariance by means of a suspicious-looking verificationist argument which, he says, "takes away from space and time the last remnants of physical objectivity."¹³ A study of his tortured path to general relativity¹⁴ shows the repeated use of this Machist line, always used to deny that some concept has a real referent. Whatever other, competing strains there were in Einstein's philosophical orientation (and there certainly were others), it would be hard to deny the importance of this instrumentalist/positivist attitude in liberating Einstein from various realist commitments. Indeed, on another occasion, I would argue in detail that without the "freedom from reality" provided by his early reverence for Mach, a central tumbler necessary to unlock the secret of special relativity would never have fallen into place.¹⁵ A few years after his work on general relativity, however, roughly around 1920, Einstein underwent a philosophical conversion, turning away from his positivist youth (he was forty-one in 1920) and becoming deeply committed to realism.¹⁶ His subsequent battle with the quantum theory, for example, was fought much more over the issue of realism than it was over the issue of causality or determinism (as it is usually portrayed). In particular, following his conversion, Einstein wanted to claim genuine reality for the central theoretical entities of the general theory, the four-dimensional space-time manifold and associated tensor fields. This is a serious business for if we grant his claim, then not only do space and time cease to be real but so do virtually all of the usual dynamical quantities.¹⁷ Thus motion, as we understand it, itself ceases to be real. The current generation of philosophers of space and time (led by Howard Stein and John Earman) have followed Einstein's lead here. But, interestingly, not only do these ideas boggle the mind of the average man in the street (like you and me), they boggle most contemporary scientific minds as well.¹⁸ That is, I believe the majority opinion among working, knowledgeable scientists is that general relativity provides a magnificent organizing tool for treating certain gravitational problems in astrophysics and cosmology. But few, I believe, give credence to the kind of realist existence and nonexistence claims that I have been mentioning. For rel-

ativistic physics, then, it appears that a nonrealist attitude was important in its development, that the founder nevertheless espoused a realist attitude to the finished product, but that most who actually use it think of the theory as a powerful instrument, rather than as expressing a "big truth."

With quantum theory, this sequence gets a twist. Heisenberg's seminal paper of 1925 is prefaced by the following abstract, announcing, in effect, his philosophical stance: "In this paper an attempt will be made to obtain bases for a quantum-theoretical mechanics based exclusively on relations between quantities observable in principle."¹⁹ In the body of the paper, Heisenberg not only rejects any reference to unobservables; he also moves away from the very idea that one should try to form any picture of a reality underlying his mechanics. To be sure, Schrödinger, the second father of quantum theory, seems originally to have had a vague picture of an underlying wavelike reality for his own equation. But he was quick to see the difficulties here and, just as quickly, although reluctantly, abandoned the attempt to interpolate any reference to reality.²⁰ These instrumentalist moves, away from a realist construal of the emerging quantum theory, were given particular force by Bohr's so-called "philosophy of complementarity"; and this nonrealist position was consolidated at the time of the famous Solvay conference, in October of 1927, and is firmly in place today. Such quantum nonrealism is part of what every graduate physicist learns and practices. It is the conceptual backdrop to all the brilliant successes in atomic, nuclear, and particle physics over the past fifty years. Physicists have learned to think about their theory in a highly nonrealist way, and doing just that has brought about the most marvelous predictive success in the history of science.

The war between Einstein, the realist, and Bohr, the nonrealist, over the interpretation of quantum theory was not, I believe, just a sideshow in physics, nor an idle intellectual exercise. It was an important endeavor undertaken by Bohr on behalf of the enterprise of physics as a progressive science. For Bohr believed (and this fear was shared by Heisenberg, Sommerfeld, Pauli, and Born—and all the major players) that Einstein's realism, if taken seriously, would block the consolidation and articulation of the new physics and, thereby, stop the progress of science. They were afraid, in particular, that Einstein's realism would lead the next generation of the brightest and best students into scientific dead ends. Alfred Landé, for example, as a graduate student, was interested in spending some time in Berlin to sound out Einstein's ideas. His supervisor was Sommerfeld, and recalling this period, Landé writes

The more pragmatic Sommerfeld . . . warned his students, one of them this writer, not to spend too much time on the hopeless task of "explaining" the quantum but rather to accept it as fundamental and help work out its consequences.²¹

The task of “explaining” the quantum, of course, is the realist program for identifying a reality underlying the formulas of the theory and thereby explaining the predictive success of the formulas as approximately true descriptions of this reality. It is this program that I have criticized in the first part of this paper, and this same program that the builders of quantum theory saw as a scientific dead end. Einstein knew perfectly well that the issue was joined right here. In the summer of 1935, he wrote to Schrödinger,

The real problem is that physics is a kind of metaphysics; physics describes ‘reality’. But we do not know what ‘reality’ is. We know it only through physical description. . . . But the Talmudic philosopher sniffs at ‘reality’, as at a frightening creature of the naive mind.²²

By avoiding the bogey of an underlying reality, the “Talmudic” originators of quantum theory seem to have set subsequent generations on precisely the right path. Those inspired by realist ambitions have produced no predictively successful physics. Neither Einstein’s conception of a unified field nor the ideas of the de Broglie group about pilot waves, nor the Bohm-inspired interest in hidden variables has made for scientific progress. To be sure, several philosophers of physics, including another Hilary Putnam, and myself, have fought a battle over the last decade to show that the quantum theory is at least consistent with some kind of underlying reality. I believe that Hilary has abandoned the cause, perhaps in part on account of the recent Bell-inequality problem over correlation experiments, a problem that van Fraassen calls “the charybdis of realism.”²³ My own recent work in the area suggests that we may still be able to keep realism afloat in this whirlpool.²⁴ But the possibility (as I still see it) for a realist account of the quantum domain should not lead us away from appreciating the historical facts of the matter.

One can hardly doubt the importance of a nonrealist attitude for the development and practically infinite success of the quantum theory. Historical counterfactuals are always tricky, but the sterility of actual realist programs in this area at least suggests that Bohr and company were right in believing that the road to scientific progress here would have been blocked by realism. The founders of quantum theory never turned on the nonrealist attitude that served them so well. Perhaps that is because the central underlying theoretical device of quantum theory, the densities of a complex-valued and infinite-dimensional wave function, are even harder to take seriously than is the four-dimensional manifold of relativity. But now, there comes a most curious twist. For just as the practitioners of relativity, I have suggested, ignore the *realist* interpretation in favor of a more pragmatic attitude toward the space-time structure, the quantum physicists would appear to make a similar reversal and to forget their non-

realist history and allegiance when it comes time to talk about new discoveries.

Thus, anyone in the business will tell you about the exciting period, in the fall of 1974, when the particle group at Brookhaven, led by Samuel Ting, discovered the J particle, just as a Stanford team at the Stanford Linear Accelerator Center (SLAC), under Burton Richter, independently found a new particle they called " ψ ". These turned out to be one and the same, the so-called ψ/J particle* (Mass 3,098 MeV, Spin 1, Resonance 67 KeV, Strangeness 0). To explain this new entity, the theoreticians were led to introduce a new kind of quark, the so-called charmed quark. The ψ/J particle is then thought to be made up out of a charmed quark and an anticharmed quark, with their respective spins aligned. But if this is correct, then there ought to be other such pairs anti-aligned, or with variable spin alignments, and these ought to make up quite new observable particles. Such predictions from the charmed-quark model have turned out to be confirmed in various experiments.

In this example, I have been intentionally a bit more descriptive in order to convey the realist feel to the way scientists speak in this area. For I want to ask whether this is a return to realism or whether, instead, it can somehow be reconciled with a fundamentally nonrealist attitude.²⁵ I believe that the nonrealist option is correct, but I will not defend that answer here, however, because its defense involves the articulation of a compelling and viable form of nonrealism; and that is the task of the third (and final) section of this paper.

■ | Nonrealism

Even if the realist happens to be a talented philosopher, I do not believe that, in his heart, he relies for his realism on the rather sophisticated form of abductive argument that I have examined and rejected in the first section of this paper, and which the history of twentieth-century physics shows to be fallacious. Rather, if his heart is like mine (and I *do* believe in a common nature), then I suggest that a more simple and homely sort of argument is what grips him. It is this, and I will put it in the first person. I certainly trust the evidence of my senses, on the whole, with regard to the existence and features of everyday objects. And I have similar confidence in the system of "check, double-check, triple-check" of scientific investigation, as well as the other safeguards built into the institutions of science. So, if the scientists tell me that there really are molecules, and atoms, and ψ/J particles and, who knows, maybe even quarks, then so be

* Often written " J/ψ " and referred to as the "gypsy" particle (for " J -psi"). For their discovery, Richter and Ting shared the Nobel Prize for physics in 1976.

it. I trust them and, thus, must accept that there really are such things, with their attendant properties and relations. Moreover, if the instrumentalist (or some other member of the species "non-realistica") comes along to say that these entities, and their attendants, are just fictions (or the like), then I see no more reason to believe him than to believe that *he* is a fiction, made up (somehow) to do a job on me; which I do not believe. It seems, then, that I had better be a realist. One can summarize this homely and compelling line as follows: it is possible to accept the evidence of one's senses and to accept, *in the same way*, the confirmed results of science only for a realist; hence, I should be one (and so should you!).

What is it to accept the evidence of one's senses and, *in the same way*, to accept confirmed scientific theories? It is to take them into one's life as true, with all that implies concerning adjusting one's behavior, practical and theoretical, to accommodate these truths. Now, of course, there are truths, and truths. Some are more central to us and our lives, some less so. I might be mistaken about anything, but were I mistaken about where I am right now, that might affect me more than would my perhaps mistaken belief in charmed quarks. Thus, it is compatible with the homely line of argument that some of the scientific beliefs that I hold are less central than some, for example, perceptual beliefs. Of course, were I deeply in the charmed-quark business, giving up that belief might be more difficult than giving up some at the perceptual level. (Thus we get the phenomenon of "seeing what you believe," as is well known to all thoughtful people.) When the homely line asks us, then, to accept the scientific results "in the same way" in which we accept the evidence of our senses, I take it that we are to accept them both as true. I take it that we are being asked not to distinguish between kinds of truth or modes of existence or the like, but only among truths themselves, in terms of centrality, degrees of belief, or such.

Let us suppose this understood. Now, do you think that Bohr, the archenemy of realism, could toe the homely line? Could Bohr, fighting for the sake of science (against Einstein's realism) have felt compelled either to give up the results of science, or else to assign to its "truths" some category different from the truths of everyday life? It seems unlikely. And thus, unless we uncharitably think Bohr inconsistent on this basic issue, we might well come to question whether there is any necessary connection moving us from accepting the results of science as true to being a realist.²⁶

Let me use the term 'antirealist' to refer to any of the many different specific enemies of realism: the idealist, the instrumentalist, the phenomenalist, the empiricist (constructive or not), the conventionalist, the constructivist, the pragmatist, and so forth. Then, it seems to me that both the realist and the antirealist must toe what I have been calling "the homely line." That is, they must both accept the certified results of science as on

par with more homely and familiarly supported claims. That is not to say that one party (or the other) cannot distinguish more from less well-confirmed claims at home or in science; nor that one cannot single out some particular mode of inference (such as inference to the best explanation) and worry over its reliability, both at home and away. It is just that one must maintain parity. Let us say, then, that both realist and antirealist accept the results of scientific investigations as 'true', on par with more homely truths. (I realize that some antirealists would rather use a different word, but no matter.) And call this acceptance of scientific truths the "core position."²⁷ What distinguishes realists from antirealists, then, is what they add onto this core position.

The antirealist may add onto the core position a particular analysis of the concept of truth, as in the pragmatic and instrumentalist and conventionalist conceptions of truth. Or the antirealist may add on a special analysis of concepts, as in idealism, constructivism, phenomenism, and in some varieties of empiricism. These addenda will then issue in a special meaning, say, for existence statements. Or the antirealist may add on certain methodological strictures, pointing a wary finger at some particular inferential tool, or constructing his own account for some particular aspects of science (e.g., explanations or laws). Typically, the antirealist will make several such additions to the core.

What then of the realist, what does he add to his core acceptance of the results of science as really true? My colleague, Charles Chastain, suggested what I think is the most graphic way of stating the answer—namely, that what the realist adds on is a desk-thumping, foot-stamping shout of "Really!" So, when the realist and antirealist agree, say, that there really are electrons and that they really carry a unit negative charge and really do have a small mass (of about 9.1×10^{-28} grams), what the realist wants to add is the emphasis that all this is really so. "There really are electrons, really!" This typical realist emphasis serves both a negative and a positive function. Negatively, it is meant to deny the additions that the antirealist would make to that core acceptance which both parties share. The realist wants to deny, for example, the phenomenistic reduction of concepts or the pragmatic conception of truth. The realist thinks that these addenda take away from the substantiality of the accepted claims to truth or existence. "No," says he, "they *really* exist, and not in just your diminished antirealist sense." Positively, the realist wants to explain the robust sense in which *he* takes these claims to truth or existence, namely, as claims about reality—what is really, really the case. The full-blown version of this involves the conception of truth as correspondence with the world, and the surrogate use of approximate truth as near-correspondence. We have already seen how these ideas of correspondence and approximate truth are supposed to explain what *makes* the truth *true* whereas, in fact, they function as mere trappings, that is, as superficial decorations that may well

attract our attention but do not compel rational belief. Like the extra "really," they are an arresting foot-thump and, logically speaking, of no more force.

It seems to me that when we contrast the realist and the antirealist in terms of what they each want to add to the core position, a third alternative emerges—and an attractive one at that. It is the core position itself, *and all by itself*. If I am correct in thinking that, at heart, the grip of realism only extends to the homely connection of everyday truths with scientific truths, and that good sense dictates our acceptance of the one on the same basis as our acceptance of the other, then the homely line makes the core position, all by itself, a compelling one, one that we ought to take to heart. Let us try to do so, and to see whether it constitutes a philosophy, and an attitude toward science, that we can live by.

The core position is neither realist nor antirealist; it mediates between the two. It would be nice to have a name for this position, but it would be a shame to appropriate another "ism" on its behalf, for then it would appear to be just one of the many contenders for ontological allegiance. I think it is not just one of that crowd but rather, as the homely line behind it suggests, it is for commonsense epistemology—the natural ontological attitude. Thus, let me introduce the acronym NOA (pronounced as in "Noah"), for *natural ontological attitude*, and, henceforth, refer to the core position under that designation.

To begin showing how NOA makes for an adequate philosophical stance toward science, let us see what it has to say about ontology. When NOA counsels us to accept the results of science as true, I take it that we are to treat truth in the usual referential way, so that a sentence (or statement) is true just in case the entities referred to stand in the referred-to relations. Thus, NOA sanctions ordinary referential semantics and commits us, via truth, to the existence of the individuals, properties, relations, processes, and so forth referred to by the scientific statements that we accept as true. Our belief in their existence will be just as strong (or weak) as our belief in the truth of the bit of science involved, and degrees of belief here, presumably, will be tutored by ordinary relations of confirmation and evidential support, subject to the usual scientific canons. In taking this referential stance, NOA is not committed to the progressivism that seems inherent in realism. For the realist, as an article of faith, sees scientific success, over the long run, as bringing us closer to the truth. His whole explanatory enterprise, using approximate truth, forces his hand in this way. But, a "NOAer" (pronounced as "knower") is not so committed. As a scientist, say, within the context of the tradition in which he works, the NOAer, of course, will believe in the existence of those entities to which his theories refer. But should the tradition change, say in the manner of the conceptual revolutions that Kuhn dubs "paradigm shifts," then nothing in NOA dictates that the change be assimilated as being progressive, that is, as a change where we learn more accurately about *the same*

things. NOA is perfectly consistent with the Kuhnian alternative, which construes such changes as wholesale changes of reference. Unlike the realist, adherents to NOA are free to examine the facts in cases of paradigm shift, and to see whether or not a convincing case for stability of reference across paradigms can be made without superimposing on these facts a realist-progressivist superstructure. I have argued elsewhere that if one makes oneself free, as NOA enables one to do, then the facts of the matter will not usually settle the case;²⁸ and that this is a good reason for thinking that cases of so-called "incommensurability" are, in fact, genuine cases where the question of stability of reference is indeterminate. NOA, I think, is the right philosophical position for such conclusions. It sanctions reference and existence claims, but it does not force the history of science into prefit molds.

So far I have managed to avoid what, for the realist, is the essential point, for what of the "external world"? How can I talk of reference and of existence claims unless I am talking about referring to things right out there in the world? And here, of course, the realist, again, wants to stamp his feet.²⁹ I think the problem that makes the realist want to stamp his feet, shouting "Really!" (and invoking the external world) has to do with the stance the realist tries to take vis-à-vis the game of science. The realist, as it were, tries to stand outside the arena watching the ongoing game and then tries to judge (from this external point of view) what the point is. It is, he says, *about* some area external to the game. The realist, I think, is fooling himself. For he cannot (really!) stand outside the arena, nor can he survey some area off the playing field and mark it out as what the game is about.

Let me try to address these two points. How are we to arrive at the judgment that, in addition to, say, having a rather small mass, electrons are objects "out there in the external world"? Certainly, we can stand off from the electron game and survey its claims, methods, predictive success, and so forth. But what stance could we take that would enable us to judge what the theory of electrons is *about*, other than agreeing that it is about electrons? It is not like matching a blueprint to a house being built, or a map route to a country road. For we are *in* the world, both physically and conceptually.³⁰ That is, *we* are among the objects of science, and the concepts and procedures that we use to make judgments of subject matter and correct application are themselves part of that same scientific world. Epistemologically, the situation is very much like the situation with regard to the justification of induction. For the problem of the external world (so-called) is how to satisfy the realist's demand that we justify the existence claims sanctioned by science (and, therefore, by NOA) as claims to the existence of entities "out there." In the case of induction, it is clear that only an inductive justification will do, and it is equally clear that no inductive justification will do at all. So too with the external world, for only ordinary scientific inferences to existence will do, and yet none of them

satisfies the demand for showing that the existent is really “out there.” I think we ought to follow Hume’s prescription on induction, with regard to the external world. There is no possibility for justifying the kind of externality that realism requires, yet it may well be that, in fact, we cannot help yearning for just such a comforting grip on reality. I shall return to this theme at the close of the paper.

If I am right, then the realist is chasing a phantom, and we cannot actually do more, with regard to existence claims, than follow scientific practice, just as NOA suggests. What then of the other challenges raised by realism? Can we find in NOA the resources for understanding scientific practice? In particular (since it was the topic of the first part of this paper), does NOA help us to understand the scientific method, such as the problems of the small handful or of conjunctions? The sticking point with the small handful was to account for why the few and narrow alternatives that we can come up with, result in successful novel predictions, and the like. The background was to keep in mind that most such narrow alternatives are not successful. I think that NOA has only this to say. If you believe that guessing based on some truths is more likely to succeed than guessing pure and simple, then if our earlier theories were in large part true and if our refinements of them conserve the true parts, then guessing on this basis has some relative likelihood of success. I think this is a weak account, but then I think the phenomenon here does not allow for anything much stronger since, for the most part, such guesswork fails. In the same way, NOA can help with the problem of conjunctions (and, more generally, with problems of logical combinations). For if two consistent theories in fact have overlapping domains (a fact, as I have just suggested, that is not so often decidable), and if the theories also have true things to say about members in the overlap, then conjoining the theories just adds to the truths of each and, thus, *may*, in conjunction, yield new truths. Where one finds other successful methodological rules, I think we will find NOA’s grip on the truth sufficient to account for the utility of the rules.

Unlike the realist, however, I would not tout NOA’s success at making science fairly intelligible as an argument in its favor, vis-à-vis realism or various antirealisms. For NOA’s accounts are available to these fellows, too, provided what they add to NOA does not negate its appeal to the truth, as does a verificationist account of truth or the realists’ longing for approximate truth. Moreover, as I made plain enough in the first section of this paper, I am sensitive to the possibility that explanatory efficacy can be achieved without the explanatory hypothesis being true. NOA may well make science seem fairly intelligible and even rational, but NOA could be quite the wrong view of science for all that. If we posit as a constraint on philosophizing about science that the scientific enterprise should come out in our philosophy as not too unintelligible or irrational, then, perhaps, we can say that NOA passes a minimal standard for a philosophy of science.

Indeed, perhaps the greatest virtue of NOA is to call attention to just how minimal an adequate philosophy of science can be. (In this respect, NOA might be compared to the minimalist movement in art.) For example, NOA helps us to see that realism differs from various antirealisms in this way: realism adds an outer direction to NOA, that is, the external world and the correspondence relation of approximate truth; antirealisms (typically) add an inner direction, that is, human-oriented reductions of truth, or concepts, or explanations (as in my opening citation from Hume). NOA suggests that the legitimate features of these additions are already contained in the presumed equal status of everyday truths with scientific ones, and in our accepting them both as *truths*. No other additions are legitimate, and none are required.

It will be apparent by now that a distinctive feature of NOA, one that separates it from similar views currently in the air, is NOA's stubborn refusal to amplify the concept of truth, by providing a theory or analysis (or even a metaphorical picture). Rather, NOA recognizes in "truth" a concept already in use and agrees to abide by the standard rules of usage. These rules involve a Davidsonian-Tarskian referential semantics, and they support a thoroughly classical logic of inference. Thus NOA respects the customary "grammar" of 'truth' (and its cognates). Likewise, NOA respects the customary epistemology, which grounds judgments of truth in perceptual judgments and various confirmation relations. As with the use of other concepts, disagreements are bound to arise over what is true (for instance, as to whether inference to the best explanation is always truth-conferring). NOA pretends to no resources for settling these disputes, for NOA takes to heart the great lesson of twentieth-century analytic and Continental philosophy, namely, that there *are* no general methodological or philosophical resources for deciding such things. The mistake common to realism and all the antirealisms alike is their commitment to the existence of such nonexistent resources. If pressed to answer the question of what, then, does it *mean* to say that something is true (or to what does the truth of so-and-so commit one), NOA will reply by pointing out the logical relations engendered by the specific claim and by focusing, then, on the concrete historical circumstances that ground that particular judgment of truth. For, after all, there *is* nothing more to say.³¹

Because of its parsimony, I think the minimalist stance represented by NOA marks a revolutionary approach to understanding science. It is, I would suggest, as profound in its own way as was the revolution in our conception of morality, when we came to see that founding morality on God and His Order was *also* neither legitimate nor necessary. Just as the typical theological moralist of the eighteenth century would feel bereft to read, say, the pages of *Ethics*, so I think the realist must feel similarly when NOA removes that "correspondence to the external world" for which he so longs. I too have regret for that lost paradise, and too often slip into the realist fantasy. I use my understanding of twentieth-century physics to

help me firm up my convictions about NOA, and I recall some words of Mach, which I offer as a comfort and as a closing. With reference to realism, Mach writes

It has arisen in the process of immeasurable time without the intentional assistance of man. It is a product of nature, and preserved by nature. Everything that philosophy has accomplished . . . is, as compared with it, but an insignificant and ephemeral product of art. The fact is, every thinker, every philosopher, the moment he is forced to abandon his one-sided intellectual occupation . . . , immediately returns [to realism].

Nor is it the purpose of these “introductory remarks” to discredit the standpoint [of realism]. The task which we have set ourselves is simply to show why and for what purpose we hold that standpoint during most of our lives, and why and for what purpose we are . . . obliged to abandon it.

These lines are taken from Mach’s *The Analysis of Sensations* (Sec. 14). I recommend that book as effective realism-therapy, a therapy that works best (as Mach suggests) when accompanied by historicophysical investigations (real versions of the breakneck history of my second section, “Realism and Progress”). For a better philosophy, however, I recommend NOA.³²

■ | Notes

1. In the final section, I call this postrealism “NOA.” Among recent views that relate to NOA, I would include Hilary Putnam’s “internal realism,” Richard Rorty’s “epistemological behaviorism,” the “semantic realism” espoused by Paul Horwich, parts of the “Mother Nature” story told by William Lycan, and the defense of common sense worked out by Joseph Pitt (as a way of reconciling W. Sellars’s manifest and scientific images). For references, see Hilary Putnam, *Meaning and the Moral Sciences* (London: Routledge and Kegan Paul, 1978); Richard Rorty, *Philosophy and the Mirror of Nature* (Princeton: Princeton University Press, 1979); Paul Horwich, “Three Forms of Realism,” *Synthese* 51 (1982): 181–201; William G. Lycan, “Epistemic Value” (preprint, 1982) [*Synthese* 64 (1985): 137–64]; and Joseph C. Pitt, *Pictures, Images and Conceptual Change* (Dordrecht: D. Reidel, 1981). The reader will note that some of the above consider their views a species of realism, whereas others consider their views antirealist. As explained below, NOA marks the divide; hence its “postrealism.”

2. Larry Laudan, “A Confutation of Convergent Realism,” [*Philosophy of Science* 48 (1981): 19–48; excerpted in this chapter].

3. Richard N. Boyd, “Scientific Realism and Naturalistic Epistemology,” in *PSA* (1980), vol. 2, ed. P. D. Asquith and R. N. Giere (E. Lansing: Philosophy of Science Association, 1981), 613–662. See also, Boyd’s article [“The Current Status of Scientific Realism,” *Erkenntnis* 19 (1983): 45–90], and further references there.

4. Hilary Putnam, “The Meaning of ‘Meaning,’” in *Language, Mind and Knowledge*, ed. K. Gunderson (Minneapolis: University of Minnesota Press, 1975), 131–

193. See also his article ["What is 'Realism'?" in *Meaning and the Moral Sciences* (London: Routledge and Kegan Paul, 1978), 18–33].

5. Bas C. van Fraassen, *The Scientific Image* (Oxford: The Clarendon Press, 1980). See especially pp. 97–101 for a discussion of the truth of explanatory theories. To see that the recent discussion of realism is joined right here, one should contrast van Fraassen with W. H. Newton-Smith, *The Rationality of Science* (London: Routledge and Kegan Paul, 1981), esp. chap. 8.

6. Nancy Cartwright's *How the Laws of Physics Lie* (Oxford: Oxford University Press, 1983) includes some marvelous essays on these issues.

7. Some realists may look for genuine support, and not just solace, in such a coherentist line. They may see in their realism a basis for general epistemology, philosophy of language, and so forth (as does Boyd, "Scientific Realism and Naturalistic Epistemology"). If they find in all this a coherent and comprehensive world view, then they might want to argue for their philosophy as Wilhelm Wien argued (in 1909) for special relativity, "What speaks for it most of all is the inner consistency which makes it possible to lay a foundation having no self-contradictions, one that applies to the totality of physical appearances." Quoted by Gerald Holton, "Einstein's Scientific Program: Formative Years" in *Some Strangeness in the Proportion*, ed. H. Woolf (Reading: Addison-Wesley, 1980), 58. Insofar as the realist moves away from the abductive defense of realism to seek support, instead, from the merits of a comprehensive philosophical system with a realist core, he marks as a failure the bulk of recent defenses of realism. Even so, he will not avoid the critique pursued in the text. For although my argument above has been directed, in particular, against the abductive strategy, it is itself based on a more general maxim, namely, that the form of argument used to support realism must be more stringent than the form of argument embedded in the very scientific practice that realism itself is supposed to ground—on pain of begging the question. Just as the abductive strategy fails because it violates this maxim, so too would the coherentist strategy, should the realist turn from one to the other. For, as we see from the words of Wien, the same coherentist line that the realist would appropriate for his own support, is part of ordinary scientific practice in framing judgments about competing theories. It is, therefore, not a line of defense available to the realist. Moreover, just as the truth-bearing status of abduction is an issue dividing realists from various nonrealists, so too is the status of coherence-based inference. Turning from abduction to coherence, therefore, still leaves the realist begging the question. Thus, when we bring out into the open the character of arguments for realism, we see quite plainly that they do not work.

In support of realism there seem to be only those "reasons of the heart" which, as Pascal says, reason does not know. Indeed, I have long felt that belief in realism involves a profound leap of faith, not at all dissimilar from the faith that animates deep religious convictions. I would welcome engagement with realists on this understanding, just as I enjoy conversation on a similar basis with my religious friends. The dialogue will proceed more fruitfully, I think, when the realists finally stop pretending to a rational support for their faith, which they do not have. Then we can all enjoy their intricate and sometimes beautiful philosophical constructions (of, e.g., knowledge, or reference, etc), even though, as nonbelievers, they may seem to us only wonder-full castles in the air.

8. I hope all readers of this essay will take this idea to heart. For in formulating the question as how to explain why the methods of science lead to instrumental success, the realist has seriously misstated the explanandum. Overwhelmingly, the results of the conscientious pursuit of scientific inquiry are failures: failed theories, failed hypotheses, failed conjectures, inaccurate measurements, incorrect estimations of parameters, fallacious causal inferences, and so forth. If explanations are appropriate here, then what requires explaining is why the very same methods produce an overwhelming background of failures and, occasionally, also a pattern of successes. The realist literature has not yet begun to address this question, much less to offer even a hint of how to answer it.

9. Of course, the realist can appropriate the devices and answers of the instrumentalist, but that would be cheating, and it would, anyway, not provide the desired support of realism per se.

10. Paul Teller has made this suggestion to me in conversation.

11. Ilkka Niiniluoto's "What Shall We Do with Verisimilitude?" *Philosophy of Science* 49 (1982): 181–197, contains interesting formal constructions for "degree of truthlikeness," and related versimilia. As conjectured above, they rely on an unspecified correspondence relation to the truth and on measures of the "distance" from the truth. Moreover, they fail to sanction that projection from some approximate truths to other, novel truths, which lies at the core of realist rationalizations.

12. See Gerald Holton, "Mach, Einstein, and the Search for Reality," in his *Thematic Origins of Scientific Thought* (Cambridge: Harvard University Press, 1973), 219–259. I have tried to work out the precise role of this positivist methodology in my "The Young Einstein and the Old Einstein," in *Essays in Memory of Imre Lakatos*, ed. R. S. Cohen et al. (Dordrecht: D. Reidel, 1976), 145–159.

13. A. Einstein et al., *The Principle of Relativity*, trans. W. Perrett and G. B. Jeffrey (New York: Dover, 1952), 117.

14. John Earman and Clark Glymour, "Lost in the Tensors," *Studies in History and Philosophy of Science* 9 (1978): 251–278. The tortuous path detailed by Earman is sketched by B. Hoffmann, *Albert Einstein, Creator and Rebel* (New York: New American Library, 1972), 116–128. A nontechnical and illuminating account is given by John Stachel, "The Genesis of General Relativity," in *Einstein Symposium Berlin*, ed. H. Nelkowski et al. (Berlin: Springer-Verlag, 1980), 428–42.

15. I have in mind the role played by the analysis of simultaneity in Einstein's path to special relativity. Despite the important study by Arthur Miller, *Albert Einstein's Special Theory of Relativity* (Reading: Addison-Wesley, 1981), and an imaginative pioneering work by John Earman (and collaborators) ["On Writing the History of Special Relativity," in *PSA 1982*, vol. 2, ed. P. Asquith and T. Nickles (East Lansing, Mich.: Philosophy of Science Association, 1983), 403–16], . . . I think the role of positivist analysis in the 1905 paper has yet to be properly understood. . . .

16. Peter Barker, "Einstein's Later Philosophy of Science," in *After Einstein*, ed. P. Barker and C. G. Shugart (Memphis: Memphis State University Press, 1981), 133–146, is a nice telling of this story. [See also Arthur Fine, *The Shaky Game* (Chicago: University of Chicago Press, 1986), ch. 6.]

17. Roger Jones in "Realism About What?" [*Philosophy of Science* 58 (1991): 185–202] explains very nicely some of the difficulties here.

18. I think the ordinary, deflationist attitude of working scientists is much like that of Steven Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (New York: Wiley, 1972).

19. See B. L. van der Waerden, *Sources of Quantum Mechanics* (New York: Dover, 1967), 261.

20. See Linda Wessels, "Schrodinger's Route to Wave Mechanics," *Studies in History and Philosophy of Science* 10 (1979): 311–340.

21. A. Landé, "Albert Einstein and the Quantum Riddle," *American Journal of Physics* 42 (1974): 460.

22. Letter to Schrodinger, June 19, 1935. See my "Einstein's Critique of Quantum Theory: The Roots and Significance of EPR," in *After Einstein* (see n. 16), 147–158, for a fuller discussion of the contents of this letter.

23. Bas van Fraassen, "The Charybdis of Realism: Epistemological Implications of Bell's Inequality," *Synthese* 52 (1982): 25–38.

24. See my "Antinomies of Entanglement: The Puzzling Case of The Tangled Statistics," *Journal of Philosophy* 79 (1982), for part of the discussion and for reference to other recent work.

25. The nonrealism that I attribute to students and practitioners of the quantum theory requires more discussion and distinguishing of cases and kinds than I have room for here. It is certainly not the all-or-nothing affair I make it appear in the text. [See Arthur Fine,] "Is Scientific Realism Compatible with Quantum Physics?" [in A. Fine, *The Shaky Game*, 151–171]. My thanks to Paul Teller and James Cushing, each of whom saw the need for more discussion here.

26. I should be a little more careful about the historical Bohr than I am in the text. For Bohr himself would seem to have wanted to truncate the homely line somewhere between the domain of chairs and tables and atoms, whose existence he plainly accepted, and that of electrons, where he seems to have thought the question of existence (and of realism, more generally) was no longer well defined. An illuminating and provocative discussion of Bohr's attitude toward realism is given by Paul Teller, "The Projection Postulate and Bohr's Interpretation of Quantum Mechanics," [in *PSA* 1980, vol. 2, ed. P. Asquith and R. Giere] pp. 201–223. Thanks, again, to Paul for helping to keep me honest.

27. In this context, for example, van Fraassen's "constructive empiricism" would prefer the concept of empirical adequacy, reserving "truth" for an (unspecified) literal interpretation and believing in that truth only among observables. It is clear, nevertheless, that constructive empiricism follows the homely line and accepts the core position. Indeed, this seems to be its primary motivating rationale. If we reread constructive empiricism in our terminology, then, we would say that it accepts the core position but adds to it a construal of truth as empirical adequacy. Thus, it is antirealist, just as suggested in the next paragraph below. I might mention here that in this classification Putnam's internal realism also comes out as antirealist. For Putnam also accepts the core position, but he would add to it a Peircean construal of truth as ideal rational acceptance. This is a mistake, which I expect

that Putnam will realize and correct in future writings. He is criticized for it, soundly I think, by Paul Horwich ("Three Forms of Realism") whose own "semantic realism" turns out, in my classification, to be neither realist nor antirealist. Indeed, Horwich's views are quite similar to what is called "NOA" below, and could easily be read as sketching the philosophy of language most compatible with NOA. Finally, the "epistemological behaviorism" espoused by Rorty is a form of antirealism that seems to me very similar to Putnam's position, but achieving the core parity between science and common sense by means of an acceptance that is neither ideal nor especially rational, at least in the normative sense. (I beg the reader's indulgence over this summary treatment of complex and important positions. I have been responding to Nancy Cartwright's request to differentiate these recent views from NOA. . . .)

28. "How To Compare Theories: Reference and Change," *Noûs* 9 (1975): 17–32.

29. In his remarks at the Greensboro conference [on realism, March 1982], my commentator, John King, suggested a compelling reason to prefer NOA over realism; namely, because NOA is less percussive! My thanks to John for this nifty idea, as well as for other comments.

30. "There is, I think, no theory-independent way to reconstruct phrases like 'really true'; the notion of match between the ontology of a theory and its 'real' counterpart in nature now seems to me illusive in principle." T. S. Kuhn, "Postscript," in *The Structure of Scientific Revolutions*, 2d ed. (Chicago: University of Chicago Press, 1970), 206. The same passage is cited for rebuttal by W. H. Newton-Smith, in *The Rationality of Science*. But the "rebuttal" sketched there in chapter 8, sections 4 and 5, not only runs afoul of the objections stated here in my first section, it also fails to provide for the required theory-independence. For Newton-Smith's explication of verisimilitude (p. 204) makes explicit reference to some unspecified background theory. (He offers either current science or the Peircean limit as candidates.) But this is not to rebut Kuhn's challenge (and mine); it is to concede its force.

31. No doubt I am optimistic, for one can always think of more to say. In particular, one could try to fashion a general, descriptive framework for codifying and classifying such answers. Perhaps there would be something to be learned from such a descriptive, semantical framework. But what I am afraid of is that this enterprise, once launched, would lead to a proliferation of frameworks not so carefully descriptive. These would take on a life of their own, each pretending to ways (better than its rivals) to settle disputes over truth claims, or their import. What we need, however, is less bad philosophy, not more. So here, I believe, silence is indeed golden.

32. My thanks to Charles Chastain, Gerald Dworkin, and Paul Teller for useful preliminary conversations about realism and its rivals, but especially to Charles—for only he, then, (mostly) agreed with me, and surely that deserves special mention. This paper was written by me, but cothought by Micky Forbes. I don't know any longer whose ideas are whose. That means that the responsibility for errors and confusions is at least half Micky's (and she is two-thirds responsible for "NOA"). Finally, I am grateful to the many people who offered comments and criticisms at the [Greensboro] conference, and subsequently. I am also grateful to the National Science Foundation for a grant in support of this research.

NOA's Ark—*Fine* for Realism

Arthur Fine says that scientific realism is dead, drowned by floods of criticism. In its place he puts the *natural ontological attitude*, NOA, pronounced 'Noah'. Fine thinks that NOA is a minimalist view which is neither realist nor antirealist. I think that NOA is a thoroughly realist view: in NOA's Ark the realist can sail happily above the floods of criticism.

NOA stems from the 'homely line' that we should accept the results of science as true in the same way that we accept the evidence of the senses as true. Fine writes:

Let us say, then, that both realist and antirealist accept the results of scientific investigations as 'true', on a par with more homely truths. . . . And call this acceptance of scientific truths the "core position". What distinguishes realists from antirealists, then, is what they add onto this core position. (Fine 1984a: 96 [1199])*

NOA is the core position all by itself, California-pure, without additives.

This is mysterious. As usually understood, the realism–antirealism issue centres precisely on the question of truth. As usually understood, realists can accept Fine's core position, but antirealists cannot. Positivists deny the existence of the 'theoretical entities' of science, and think that any theory which asserts the existence of such entities is *false*. Instrumentalists think that scientific theories are tools or rules which are *neither true nor false*. Epistemological antirealists like van Fraassen or Laudan concede that theories have truth-values, even that some of them might be true, but insist that no theory should be *accepted as true*. None of these antirealist positions, as usually understood, is consistent with Fine's core position.

FROM *Philosophical Quarterly* 39 (1989): 383–98.

* Page references to the Fine reading in this chapter are enclosed in brackets.

The mystery unravels, it seems, when Fine says that antirealists often go in for some peculiar theory of truth:

The antirealist may add onto the core position a particular analysis of the concept of truth, as in the pragmatic and instrumentalist and conventionalist conceptions of truth. . . . These addenda will then issue in a special meaning, say, for existence statements. (Fine 1984a: 97 [1999])

So the positivist is seen as saying, not that theoretical science is all false because there are no 'theoretical entities', but that some theoretical science is true because in its application to science 'true' *means useful*. The instrumentalist is seen as saying something similar. Van Fraassen is seen as saying, not that we should never accept a theory about the 'unobservable' as true, but that we sometimes may because in its application to such theories 'true' *means empirically adequate*. Laudan is seen as saying something similar.

It is certainly possible to see the antirealisms this way. But it is not the way the antirealists see themselves, nor is it the clearest way to see them. Adding to the core position a peculiar truth-theory for science actually demolishes that position. The results of science are not being accepted as true on a par with more homely truths. Homely statements are accepted as true in the homely sense of the term 'true'—bits of science are accepted as true in some esoteric sense of the term 'true'. The latter acceptances are not on a par with the former at all. Only the equivocation on the term 'true' could make us think otherwise.

Fine might object that there is no such equivocation: esoteric antirealist truth-theories are meant to apply both to homely truths and to scientific ones. I doubt that this will work. Some scientific theories are true, meaning useful for 'saving the phenomena'—and some statements of the phenomena are true, meaning useful for saving . . . *what?* Some statements about unobservables are true, meaning they yield nothing but truths about observables—and some statements about observables are true, meaning they yield nothing but truths about . . . *what?* The esoteric truth-theories developed for science seem to be parasitic upon the homely conception of truth being applied to homely statements about the phenomena or about observables.

But even if an esoteric truth-theory can be applied across the board, this does not help. Now realists and antirealists cannot both accept the core position, because there is no one core position for them both to accept. Now we have *several different core positions*, depending upon the meaning to be attached (across the board) to the term 'true'. Confusingly, these different positions are all expressed using the same *words*. Realists and antirealists can both assent to the *words*—but only because they mean quite different things by them.

I wrote a paragraph back about 'the homely conception of truth being

applied to homely statements'. It may be objected that there is no such 'homely conception of truth', that the core position leaves it open what sense is to be attached to the term 'true', that antirealists attach esoteric senses to the term, *and so do realists*. This does not help either—nor does it seem to be Fine's own view of the matter.

It does not help because now we do not have different core positions confusingly expressed by the same words, but *no core position at all*. With no sense yet attached to the term 'true', neither the realist nor the anti-realist knows what it is to assent to the so-called 'core position'. If each assents to it by tacitly reading 'true' a different way, then we are back to the plethora of core positions.

Nor, it seems, does Fine think that the core position (NOA) leaves it open what sense is to be attached to the term 'true'. On the contrary, a very definite conception of truth is built into it:

When NOA counsels us to accept the results of science as true, I take it that we are to treat truth in the usual referential way, so that a sentence (or statement) is true just in case the entities referred to stand in the referred-to relations. Thus, NOA sanctions ordinary referential semantics, and commits us, via truth, to the existence of the individuals, properties, relations, processes, and so forth referred to by the scientific statements that we accept as true. (Fine 1984a: 98 [1200])

NOA recognizes in "truth" a concept already in use and agrees to abide by the standard rules of usage. These rules involve a Davidsonian-Tarskian referential semantics, and they support a thoroughly classical logic of inference. Thus NOA respects the customary "grammar" of 'truth' (and its cognates). (Fine 1984a: 101 [1203])

Now 'Davidsonian-Tarskian referential semantics' is already a philosophical theory or analysis of truth. It is hardly part of the 'standard rules of usage' for the term 'true', though it yields or explains many of those rules. No matter. The key point is that referential semantics yields precisely the notion of truth that *realists* want to apply across the board, both to homely truths and to scientific ones. Antirealists of any ilk could not accept the core position once they realized that this committed them to accepting some scientific theories as true *in the usual referential way*. NOA, the core position all by itself, is already a thoroughly realist position.

Fine thinks otherwise. But he has some difficulty in explaining what the realist adds to the core position, what distinguishes the NOA (pronounced 'knower') from the realist. It transpires that the realist shouts while the NOA speaks quietly, that the realist trafficks in certain slogans that the NOA avoids, and that the realist has a certain metaphysical picture that the NOA does not have. Let us consider these one by one, beginning with the shouting. Fine writes:

What, then, of the realist, what does he add to his core acceptance of the results of science as really true? . . . what the realist adds on is a desk-thumping, foot-stamping shout of "Really!" (Fine 1984a: 97 [1199])

So the NOA is a realist who avoids desk-thumping, foot-stamping, and shouting. Whereas older realists shouted and stamped in opposition to antirealist conceptions of truth, on NOA's Ark realists content themselves with a 'stubborn refusal to amplify' their referential semantic concept of truth. As NOA's Ark sails into the sunset, it carries only polite and restrained realists. I promise to shout no more, so that I can join this happy ship.

The slogans will take longer to dispose of. The realist is supposed to traffick in slogans, like 'Truth is correspondence with Reality', which the NOA avoids. Now for me the content of this slogan is exhausted by something the NOA does say: 'A sentence (or statement) is true just in case the entities referred to stand in the referred-to relations'. I have always thought (with Tarski himself) that the semantic conception of truth is a version of the common-sense correspondence theory of truth. Earlier correspondence theorists gave partial accounts (like Aristotle's) or trafficked in general slogans (like the one under discussion). Tarski (1944) showed how to give a complete account (for each well-defined language) which explains what the slogans mean. In this way Tarski rehabilitates the common-sense correspondence theory.

Most philosophers (including Arthur Fine) think that the correspondence theory of truth is one thing and Tarski's theory another. What does the correspondence theorist provide (or seek to provide) which Tarski does not? Fine thinks he provides (or seeks to provide) a general *account* of correspondence, a *theory* of the way in which language and reality can 'match up' so that truth results. Such a theory would 'explain what *makes* the truth true' (Fine 1984a: 97 [1199]). Armed with an account of *the* relation which all truths bear to reality, the correspondence theorist will know what all truths have in common, their 'essence', what makes them a 'natural kind'. And here Fine is sceptical:

Of course we are all committed to there being some kind of truth. But need we take that to be something like a "natural" kind? (Fine 1984b: 56)

I share the scepticism. One thing Tarski taught us is that an *essentialist* correspondence theory is out, that the *essence* of truth is a chimera. (If we must talk of 'essences' here, let us talk thus: Tarski gives us the essence of the correspondence theory *without* giving truth an essence.) Truths are many and various, and so are the ways they correspond to facts. Tarskian referential semantics captures 'the' correspondence relation as well as it *can* be captured. Michael Levin puts it well:

Tarski tells us that all true conjunctions have in common the truth of each conjunct, that each true existential generalisation is such that its matrix is satisfied by at least one sequence, and so on. To be sure, at the level of the basis clauses, the definition goes strongly extensional, but that is the way it ought to go. The truths "Ron Reagan is a man" and "This tulip is red" are not shown to have much in common: each is a matter of a different object . . . satisfying a different primitive open sentence. But *do* they have more in common than being a man has in common with being red, which is to say, very little? I cannot see that they do, or at least more than Tarski gives them. (Levin 1984: 126)

When you think about it, Tarski's work shows that the idea that truths might all share an essence is quite absurd. It is meaningful *linguistic* items that are true or false in Tarski's sense. Languages are largely conventional human inventions, suited to different human purposes. Why suppose that the bewildering variety of truths languages contain form a *natural kind*? It is surely better to suppose that there is no more to the 'correspondence relation' than Tarski gives us. Nor need it be any part of scientific realism to think that there is more.

Fine disagrees. He thinks realists must add to Tarskian truth a hankering after truth's essence. In this connection he mentions Hilary Putnam. Now with friends like Putnam, realism needs no enemies. Perhaps Putnam did hanker after truth's essence—but he did not find it. Instead, he found his model-theoretical argument against realism, and abandoned the realist cause. (By the way, the model-theoretic argument is invalid—but that would be another paper.) Never mind Putnam. Realists *need* not hanker after truth's essence (PROOF: *I don't*). Realists can think that Tarski gives them as much of a correspondence theory as they need. But perhaps, in view of all the dust philosophers have raised with the word 'correspondence' down the ages (only to complain afterwards that they cannot see), realists would do better to drop the word. Perhaps realists would do better to drop the correspondence slogan too in favour of that of the NOA: 'A sentence (or statement) is true just in case the entities referred to stand in the referred-to relations'. [1200]

So much for one philosophical worry about Tarski's theory of truth: whether or to what extent it is a *correspondence* theory of truth. There are many other philosophical worries, and I want to digress into one of them, for it will reveal another possible way to drive a wedge between NOA and scientific realism.

I have been arguing that Tarski's theory gives the scientific realist all he needs from a theory of truth. But since Tarski's theory can be applied across the board, so to speak, will it not give all sorts of *suspect* realists all that they need also? That the theory applies across the board can be seen from the following list of instances of Tarski's *Convention T*:

- 1 The statement 'There is a full moon tonight' is true if and only if there is a full moon tonight.
- 2 The statement 'Electrons are negatively charged' is true if and only if electrons are negatively charged.
- 3 The statement 'Two plus two equals four' is true if and only if two plus two equals four.
- 4 The statement 'Eating people is wrong' is true if and only if eating people is wrong.
- 5 The statement 'Ronald Reagan gives me the creeps' is true if and only if Ronald Reagan gives me the creeps.

The worry is this. Suppose that (1) yields common-sense realism about the moon, and (2) yields scientific realism about electrons. Then will not (3) yield *Platonic realism* about natural numbers, (4) *moral realism* about wrongness and rightness, and (5) realism regarding a mysterious entity (*the creeps*) which Ronald Reagan gives to me (and simultaneously, no doubt, to others too)? Since creeps-realism is absurd, and moral realism and Platonism philosophically suspect, so is the Tarskian theory which yields them.

This worry is quite groundless. Tarski's Convention T (often called Tarski's 'disquotational scheme', or a disquotational, deflationary, or redundancy theory of truth) *by itself yields none of these realisms*. We all avoid creeps-realism by saying that 'Ronald Reagan gives me the creeps', though true (which it is), is an idiom which is not to be taken at face value for logico-philosophical purposes. We replace it with a non-idiom (say, 'Ronald Reagan makes me nervous') and avoid ontological commitment to *the creeps*. Similarly, one sceptical of moral realism might refuse to take 'Eating people is wrong' at face value for logico-philosophical purposes—which is just what emotivists, prescriptivists, and the like do. Alternatively, one might go in morals the way Hartry Field (1982) goes in mathematics. Field accepts Tarski's scheme *and* takes 'Two plus two equals four' at face value, but eschews arithmetical Platonism by saying that it is false *because* there are no numbers for the numerals 'two' and 'four' to be names of. Moral realism might be eschewed similarly: take 'Eating people is wrong' at face value, apply Tarski's scheme to it, and say that it is false *because* there is no property for 'wrong' to refer to.

Similarly again, those sceptical of scientific realism can *either* refuse to take 'Electrons are negatively charged' at face value (as instrumentalists do in their different ways), *or* take it at face value and say that it is false because there are no theoretical entities for the term 'electron' to refer to (as positivists do), *or* take it at face value and concede that it might be true but insist that we should not accept it as true (as epistemological anti-realists do). Finally, those sceptical of common-sense realism can *either* refuse to take 'There is a full moon tonight' at face value, *or* take it at face value and say that it is false because there is no external object for the

word 'moon' to refer to (as Bishop Berkeley and the phenomenologists do). The disjunctions here are not, of course, exclusive ones. Indeed, antirealists typically adopt a mixed strategy, first saying that a certain kind of statement is false taken at face value (or taken literally), then trying to soften that conclusion by telling us what such statements 'really mean', that is, how they are to be taken for logico-philosophical purposes. (Think of phenomenalist 'translations', so called, of external-object statements, or of emotivist 'translations', so called, of moral statements.) The idea that Tarski's Convention T *by itself* begs all kinds of metaphysical questions is simply mistaken. Tarski himself said as much long ago.

By the way, calling Convention T a 'disquotational theory of truth' is highly misleading. Tarski's theory of truth is not exhausted by Convention T, though most of the philosophical worries about it are. Instances of Convention T will not be 'disquotational' if something other than a quotation-mark name of a statement appears on the left-hand side, or if the statement talked about comes from a different language from the one in which we talk. It is cases where neither of these conditions obtains, cases like my (1)–(5), that are also responsible for the mistaken view that Convention T is trivial *because circular*. There are similar reservations, for similar reasons, about calling the theory a 'deflationary' theory or a 'redundancy' theory.

If Tarski's Convention T does not by itself yield realism about the moon, electrons, natural numbers, moral properties, the creeps, or anything else, *what is its importance for realism?* The answer is obvious: it makes realism about all of these things *possible*. This is its importance for realism; this is why it is the theory of truth that realists need. Antirealist theories of truth identify it with some *internal* feature of our beliefs (their coherence, their usefulness, their self-evidence, their ultimate undisbelievability, or whatever).¹ Such theories make realism impossible; they leave no room for it. So realists need Tarski's theory of truth—but they also need more. To be a *realist about Xs* (whatever Xs may be) you must:

- a take statements about Xs at face value for logico-philosophical purposes;
- b apply Tarski's Convention T to those statements;
- c accept some of those statements (*appropriate* ones: 'There are no Xs' will not do) as true.

Let us end the digression and return to Fine's NOA. He does all of these three things when it comes to the theoretical statements of science. He insists that such statements are to be taken at face value (unlike the instrumentalists). He applies Tarski's scheme to them (unlike those who traffick in unrealistic truth-theories). And he accepts appropriate theoretical statements as true (unlike the positivists or epistemological antirealists). All this places Fine's NOA squarely in the realist camp.

But wait! The digression may have a Fine point after all. It reveals a possible way to drive a wedge between the NOA and the realist. Suppose, contrary to what was just said, that the NOA does *not* take at face value any body of statements about Xs, whatever X might be. Suppose that to do this is already to *add* to the California-pure core position. The NOA is philosophically neutral between realist and instrumentalist interpretations of theoretical scientific statements like (2). The NOA is also neutral between realist and phenomenalist interpretations of common-sense statements like (1). Bishop Berkeley counts as a NOA—even a solipsist counts as a NOA. The NOA accepts homely statements and scientific statements as true. The NOA reads ‘true’ Tarski-style, with all this brings in the way of commitment to the entities referred to in accepted statements which are taken at face value. *But* the NOA leaves it open what those commitments actually are, because he leaves it open which statements are to be taken at face value (that is, realistically) and which are not.

This is a possible position. It is consistent (just) with Fine’s accounts of NOA. It attributes to the NOA (pronounced ‘knower’, remember) *a complete philosophical know-nothing-ism*. The NOA is not committed to electrons, the moon, tables and chairs, physical objects, other people, his self; *anything at all*. I first overlooked this possible interpretation of Fine’s NOA. It was suggested to me by discussions with Arthur Fine in Indiana in 1987 (discussions for which I am grateful). I now think it might be the correct interpretation—as our unfinished business will show.

The NOA was said to differ from the realist on three counts. We have dealt with two of these: realists need not shout or stamp; and they can confine themselves to slogans acceptable to the NOA, like ‘A sentence (or statement) is true just in case the entities referred to stand in the referred-to relations’. I cite this slogan for the third time because it would seem to give the realist all he needs by way of *metaphysics*. Fine disagrees, which brings us to the third alleged point of difference: the realist has a metaphysical picture which the NOA lacks.

Fine describes the realist metaphysic thus:

For realism, science is *about* something; something *out there*, ‘external’ and (largely) independent of us. The traditional conjunction of externality and independence leads to the realist picture of an objective, external world; what I shall call the *World*. According to realism, science is about *that*. Being about the *World* is what gives significance to science. (Fine 1986a: 150)

What exactly is wrong with this realist metaphysical picture? Is not the NOA committed to precisely the same picture?

Fine talks about ‘the obscurity of the correspondence relation and the inscrutability of realist-style reference’. He elaborates:

The problem is one of access. The correspondence relation would map true statements (let us say) to states of affairs (let us say). But if we want to compare a statement with its corresponding state of affairs, how do we proceed? How do we get at a state of affairs when that is to be understood, realist-style, as a feature of the *World*? (Fine 1986a: 151)

What exactly is the problem here? Somebody says "There is a full moon tonight", and I look up into the night sky and ascertain that the statement is true. (I use humdrum commonsensical examples, rather than esoteric scientific ones, because if there is a problem, it will be a quite general one, which afflicts the common-sense realist metaphysic as much as the scientific one.) I have access to both terms of the 'correspondence relation': my linguistic competence gives me access to what was said, my eyes give me access to the full moon out there in the world (or if you prefer, the *World*). Of course, *explaining* linguistic competence or sensory awareness gives rise to a host of scientific and philosophical problems—but to explain them is not to explain them away, to show that they do not exist at all.

Perhaps the worry is that if I were to report pedantically (similarity to (1) is intended) 'The statement "There is a full moon tonight" is true since there is a full moon tonight', then I would still be trapped inside language, would not have got 'at a state of affairs . . . understood, realist-style, as a feature of the *World*'. Perhaps the worry is that instances of Convention T, such as (1), do not relate language to the World, but rather one language (that talked about) to another (that in which we talk).

If this is the worry, it is a queer one. (1) speaks about a bit of language *and* about the World. True, to speak about the way in which language relates to the World, one must *use language*. But this is no deep truth; rather, it is a pallid truism. Sweeney had it right:

. . . I gotta use words when I talk to you
But if you understand or if you don't
That's nothing to me and nothing to you
We gotta do what we gotta do . . .*

We are not trapped inside language in the *serious* sense that all we ever talk *about* is language. To think otherwise is to ignore the hard-won distinction between use and mention.^{2†}

* From T. S. Eliot's unfinished poem, "Sweeney Agonistes: Fragments of an Aristophanic Melodrama," in *Collected Poems 1909–1962* (New York: Harcourt, Brace and World, 1963), 123.

† The sentence, "April has five letters," is about the word *April*, not the month it names; whereas "April is the cruellest month" is about the month, not its English name. The first sentence mentions the word *April* to talk about the name; the second sentence uses the word *April* to talk about the month.

I hesitate to attribute this worry to Arthur Fine. So let us see how he continues:

A similar question comes up if we move to reference and try to establish truth-conditions compositionally, for there again, what the realist needs by way of the referent for a term is some entity in the *World*. The difficulty is that whatever we observe, or, more generously, whatever we causally interact with, is certainly not independent of us. This is the problem of *reciprocity*. Moreover, whatever information we retrieve from such interaction is, *prima facie*, information about interacted-with things. This is the problem of *contamination*. How then, faced with reciprocity and contamination, can one get entities both independent and objective? Clearly the realist has no direct access to his *World*. (Fine 1986a: 151)

Return to my humdrum example (once again, any problems here will afflict humdrum examples as much as esoteric ones). There the term 'moon' referred to the moon, which is an entity out there in the *World* if anything is. Do the 'problems' of reciprocity and contamination show that this is incorrect? Reciprocity is supposed to show that the moon is not independent of us because we can see it or otherwise causally interact with it. But implicit in this is a silly account of independence: an entity is independent of us if we cannot causally interact with it. The only independent entities in this sense will be Platonic entities, which do not exist in space and time, and which have no relations causal or otherwise with beings (like us) which do exist in space and time. No scientific realist should accept an account of independence which means that the only independent entities are Platonic entities and the only independent reality the Platonic realm of abstract entities. When a scientific realist says that the moon is (largely) independent of us, he obviously means that it is nonmental, it exists outside of us, we did not create it, it existed long before we did, it continues to exist when we are not looking at it, and so forth.

What of the 'problem' of *contamination*? It is supposed to show that when we see that the moon is full, we gain information not about an objective moon out there in the *World*, but rather about an *interacted-with-moon*. This hyphenated entity is presumably not the same entity as the moon (or, if hyphens thrill you, as the moon-in-itself). Fine suggests that, unlike the moon-in-itself, the interacted-with-moon is not objective, not out there in the *World*. Where is it then, subjective and inside our heads? This smacks of the long-discarded view that we do not see external objects like the moon at all, but rather moonish-sense-data inside our heads. I doubt that Fine wants to return to that view. I know that Fine's NOA ought not to attach that bit of bad philosophy to his core acceptance of the results of science.

Perhaps the thought is that the interacted-with-moon, the moon-as-

observed-by-us, is not 'objective' in the sense that it is somehow partly constituted by the moon-concept which is our invention. After all, in order to see that there is a full moon, I must possess the moon-concept, and in order to say that there is a full moon, I must possess the word 'moon'. What I can see (or say) depends partly upon the concepts (words) that I possess.³ The world (the world-in-itself, Fine's World-with-a-capital-'W') is not carved up according to any conceptual or linguistic scheme. It is we who carve things up according to such schemes. Having carved, we cannot partake of the world-in-itself, the world-as-it-is-independently-of-any-conceptual-scheme, the World-with-a-capital-'W'. That the moon is full is not a fact about the world-in-itself, since it trafficks in the moon-concept. The statement 'The moon is full tonight' does not state a truth about the World-with-a-capital-'W', since it trafficks in the word 'moon'. The world we experience and talk about is not the world-as-it-is-independently-of-any-conceptual-scheme. Rather, it is a world partly of our own conceptual or linguistic making, a world-as-conceived-by-us or a world-as-talked-about-by-us. This is conceptual (or linguistic) *idealism*.

It quickly turns into conceptual (or linguistic) *relativism*. Our concepts vary and change. There is no *one* world-as-conceived-by-us at all. The world-as-conceived-by-the-Aristotelian differs radically from the world-as-conceived-by-the-Newtonian. The world-of-the-Eskimo is not at all the world-of-the-Kalahari-bushman. This gets really exciting once we cease to be human chauvinists and consider non-human animals too. They experience the world too, but differently from us. The world-of-the-chimpanzee is not at all the world-of-Albert-Einstein, and both are *worlds apart* from the world-of-the-honeybee or the world-of-the-three-spined-stickleback. And so on and so forth—tediously.

Kant is, of course, the philosopher who started the rot here. Kant stopped the rot from spreading, blocked the slide from idealism into relativism, by assuming that humans all have the same immutable set of basic concepts. Contemporary philosophical wisdom has outgrown that assumption. Even if contemporary wisdom is misguided, we might still ask Kant and the Kantians about non-human animals. Do they *experience* the world at all? If so, do they possess all those Kantian categories of the understanding deployment of which is a condition of the possibility of all experience? Do chimps, honeybees, and flatworms structure incoming stimuli the way humans do? One alternative is to say that humans are the only animals that have experiences. This must be deemed implausible by anyone who takes Darwinism seriously. Another alternative is to say that the Kantian categories are only conditions of the possibility of *human* experience and that other creatures can do without them. But if chimps can do without them, why not us?

Of course, all this talk of different worlds-as-experienced (conceived, talked about)-by-Xs *need not be taken seriously*. We can see it just as a fancy way of drawing attention to the great diversity of experience, con-

cepts, and talk of *the world*. On this view, all such entities as the moon-as-experienced-by-us are ersatz entities. (After all, is not 'ersatz' German for 'hyphenated', and 'hyphenated' philosopher's English for 'artificial' or 'unnatural' or 'unreal'?) The moon-as-experienced-by-us is just the moon—and similarly for all other hyphenated entities (*including* the Kantian moon-in-itself). On this view, 'The moon-as-conceived-by-Aristotelians was perfectly spherical' is just philosopher's gobbledy-gook for 'Aristotelians thought the moon is perfectly spherical'.

If we do take talk of different worlds-as-experienced (conceived, talked about)-by-Xs seriously, we become experiential (conceptual, linguistic) idealists. And we come to inhabit a strange world indeed. Consider the moon-as-observed-by-us (*moon*₀ for short) and the moon-in-itself (*moon*₁ for short). Is *moon*₀ identical with *moon*₁? Presumably not: if it were, the distinction would have no point, and we could rest content just with the *moon* (without subscript, unhyphenated). But if *moon*₀ is distinct from *moon*₁, then there is presumably some property which the one lacks and the other possesses. But to know this is to know something about *moon*₁, when our knowledge was supposed to be confined to *moon*₀! Certainly, there could be no empirical evidence that *moon*₀ is different from *moon*₁—it is just a piece of idealist metaphysics. Kantians object that we can know, not through evidence but through transcendental argument (whatever that is), that *moon*₁ not only *exists*, but also lacks various properties that *moon*₀ possesses. For example, *moon*₁ has no position in space and time, these being 'forms of sensibility' in which only *moon*₀ is located. Nor does *moon*₁ cause (or help cause) moon-visions down on earth, causality being a category of the understanding which applies only in the world-of-appearance, not in the world-of-things-in-themselves. No wonder that some of Kant's immediate followers, realizing that *moon*₁ is nowhere, at no time, and does nothing, decided that it was an *idle* metaphysical posit, did away with it altogether, and became fully-fledged idealists.* At this point my

* Kant thought that we possess synthetic a priori knowledge about time, space, geometry, and so on—for example, that we know a priori that all external objects are spatial and are causally related to one another—but that we have this knowledge only insofar as the "objects" in question are objects-as-experienced-by-us not things as they are in themselves, independent of human experience. Because things-in-themselves are not subject to the categories of causation, substance, space, time, and so on, some later thinkers (notably, Hegel) dispensed with things-in-themselves entirely, thus precipitating what Musgrave judges to be the disastrous descent into full-fledged idealism. For the idealist, the entire external world is simply a construction of the human mind (or the mind of God) and cannot exist independently of mind. By retaining things-in-themselves in his system, Kant sought to avoid this extreme sort of idealism. For Kant, although things-in-themselves cannot be said to be spatial, or to exist in time, or to cause our perceptions, they nonetheless exist independently of human minds and are the ultimate source (in some noncausal sense) of our perceptions. See Anthony Quinton, "The Trouble with Kant," *Philosophy* 72 (1997): 5–18.

Kantian friends (and I still have one or two) tell me that Kant was an 'empirical realist' and only a 'transcendental idealist'. I do not understand these Kantian slogans. I am reminded how fond Berkeley was of presenting himself as a defender of common-sense realism. The Kantian metaphysic, seen as it really is, is a form of *idealism*, as is Berkeley's metaphysic. Modern idealism is just Kantian idealism 'gone linguistic' or 'gone conceptual' and generalized.⁴

We have come a long way from Fine's remark that when we observe or otherwise interact with things, the information we retrieve is information about interacted-with-things. The remark may have been a perfectly innocent one. I do not know whether Fine is an idealist of this kind (or rather, of these kinds). I do think that Fine's NOA should have nothing to do with these idealisms. NOA stands for *natural* ontological attitude. The ersatz hyphenated entities involved in these idealisms are artificial and unnatural entities. These idealisms are dubious, perhaps in the end unintelligible, philosophical theories which no NOA should attach to his core acceptance of the results of science. Indeed, some pretty mundane and well-entrenched results of science tell us that the moon (not some hyphenated moon, not even the Kantian moon-in-itself, just the moon) is objective and independent of us: it exists outside of our heads, it was not created by us, it existed long before we did, and so forth. The NOA who accepts these bits of science as true has precisely the same metaphysical picture as the realist. Fine rejects the realist's metaphysical picture, not as unproved, but as false. Its falsity follows from what might be called an unnatural idealist attachment to science. But the unphilosophical NOA ought not to be trafficking in the 'problems' of reciprocity and contamination—for that traffic is *philosophy*. The unphilosophical NOA should do no more than accept homely truths and scientific ones. Will that not provide the NOA with the same 'metaphysical picture' as the realist?

An affirmative answer to this question overlooks the possibility of the NOA who knows *nothing* philosophical. That NOA's acceptance of bits of science as true implies nothing whatever about the objectivity and independence of the moon. For that NOA leaves it quite open how the accepted statements are to be 'interpreted', what they mean, what their ontological commitments are. Perhaps homely truths and scientific ones are to be taken at face value for logico-philosophical purposes, and *perhaps they are not*. Perhaps the homely truths are to be given a phenomenalist construal and the scientific truths an instrumentalist one. (Remember, Berkeley counts as a NOA in this minimalist sense.) The completely unphilosophical NOA leaves all this open.

In traditional discussions of scientific realism, common-sense realism regarding tables and chairs (or the moon) is accepted as unproblematic by both sides. Attention is focused on the difficulties of scientific realism regarding 'unobservables' like electrons. But Fine's discussion is not a traditional discussion. Fine's NOA, on this interpretation, begs *no* metaphys-

ical question. That is why both realists and antirealists of any ilk (even a solipsist) can accept NOA.

One special consideration suggests that this is the correct interpretation of Fine's position. In his acclaimed biography of Einstein, Abraham Pais tells how Einstein once turned to him and asked 'if I really believed that the moon exists only if I look at it' (Pais 1982: 5). Einstein's question concerned, of course, that interpretation of quantum mechanics according to which entities do not exist in well-defined states unless they are being observed. What if quantum mechanics, thus interpreted, should turn out to be correct? Will not science have turned out to overthrow even common-sense realism and to vindicate Bishop Berkeley? Science and common sense have often clashed—why should the clash not be as radical as this one? If this is possible, then we should not encumber our *philosophy* of science with metaphysical assumptions (those of common-sense realism) which science may outgrow. Hence the minimalist position called NOA is the only defensible position in *philosophy* of science—the rest is *up to science*.

Let me hasten to add that Fine himself is no friend of the interpretation of quantum mechanics just mentioned. Indeed, he has taken Einstein's side in his discussions of those debates (particularly in his 1986b). The point is that he wishes to define a philosophy of science (NOA) which will leave open all metaphysical questions for science to decide.

I doubt that this will work. The interpretation of quantum mechanics just mentioned is not a result of *science*—it is a philosophical interpretation of science inspired in some part by dubious philosophical theories like the verifiability theory of meaning. Science unaided by philosophy could not overthrow common-sense realism. Indeed, the quantum physicist presupposes common-sense realism every time he sets up some experimental apparatus (as the principle of complementarity acknowledges). Most of the alleged clashes between science and common sense (but not all of them) stem from Eddington's (1928) mistake (to *explain* the solidity of the table, in terms of the behaviour of things that are not solid, is to *explain solidity away*).*

But these are difficult questions, not to be answered in a paragraph.

* As noted in the first paragraph of Ernest Nagel, "Issues in the Logic of Reductive Explanations," (in chapter 8), Eddington refers to his "two tables" in the introduction to his Gifford Lectures, published as *The Nature of the Physical World* (New York: Macmillan, 1928). The first table is the colored, solid, wooden piece of furniture at which Eddington is seated, writing his lectures. The second is what Eddington calls his "scientific" table, which is mostly empty space occupied by electrons and other charged particles moving around at great speed. Eddington writes that "modern physics has by delicate test and remorseless logic assured me that my second scientific table is the only one which is really there—wherever 'there' may be" (xii). According to Eddington, the "external world of physics [is] a world of shadows" which is "all symbolic" (xiv).

Let us return from them to the minimalist position NOA, to see how minimalist it really is. The NOA is supposed to 'accept', say, 'There is a full moon tonight' and 'Electrons are negatively charged'. But does the NOA know what he has accepted? Remember, he leaves it open whether or not these statements are to be taken at face value, says nothing about how they are to be interpreted, what their ontological commitments are. The unphilosophical NOA does not just know nothing philosophical—he knows nothing at all.

We need to inject some content into NOA (pronounced 'knower'). One way to do this is clear: let the NOA give 'homely truths' a face-value realist construal. (Scientific realists and their opponents were agreed upon this much anyway.) But then the same will apply to the scientific statements which the NOA accepts 'on a par with' the homely ones. And NOA's Ark will after all be Fine (with a capital 'F') for realism (with a small 'r').

■ | Notes

1. Down the ages the chief motive for antirealist truth-theories has been anti-sceptical, to make the truth accessible to us. By identifying truth with some internal features of beliefs, they make it something the believer is an authority on and can know for certain. Hence they are all aptly called *subjective* truth-theories.

Their immediate consequence is *relativism* about truth. If A's belief that *P* possesses the internal feature and B's belief that *P* does not, then *P* is true for A and false for B. (Alternatively put, the 'laws of truth', contradiction and excluded middle, fail.)

To avoid relativism (and preserve the laws of truth), subjective truth-theorists tend to go *ideal* and to the *long run*. For example, coherence theorists will say that something is true (by definition) if it 'coheres', not with your or my beliefs, but with the beliefs an *ideally rational inquirer* would have in the long run.

Of course, such moves immediately threaten the anti-sceptical virtues (if virtues they be) of subjective truth-theories. What God will be coherently believing at the end of time is just as inaccessible to you or me as is truth-as-correspondence. (I think it is *more* inaccessible.)

To solve these problems, subjectivists tend to adopt a policy of *flipping back and forth*. When epistemological concerns are paramount, they stay subjective; when semantic concerns are paramount, they go ideal and to the long run. The resulting fandango is one of the least edifying sights in philosophy. Let us view it no more. (Fine's [1984b] has some excellent arguments against antirealist truth-theories.)

2. Wittgenstein ignored this distinction in his *Tractatus*, and was led to the following 'deep' thoughts: the way language relates to the world cannot be said, it can only be shown; the limits of my language are the limits of my world; what the solipsist means is quite correct, only it cannot be said. Wittgenstein's 'logocentric predicament' is simply old psychologistic wine poured into new linguistic

bottles. The British empiricists thought that thinking consists in having a stream of 'ideas', and concluded mistakenly that all we ever think *about* are our own ideas.

3. Between this sentence and its predecessor there is a subtle slide which I have not disrupted the text to remark upon. It is trivially true that a being lacking the typewriter-concept (the word 'typewriter') cannot *see that* there is a typewriter on the table. It is trivially false that a being lacking the typewriter-concept (the word 'typewriter') cannot *see* the typewriter on the table. The Kalahari bushman may see the typewriter perfectly well, as evidenced by his response to the request (couched in Kalahari-bushman-ese) to pass him that thing over there. The cat might see the typewriter perfectly well, as evidenced by her not bumping into it when the mouse she is chasing hides under it. The slide results from conflating *seeing-that* with *seeing*. (In between these there is *seeing-as*: the cat or the Kalahari bushman may see the typewriter as non-food, and the mouse as food, even though they lack the typewriter-concept or the mouse-concept.)

This slide is responsible for the view that beings possessed of different 'conceptual schemes', different languages, even different theories, literally see different worlds. The Aristotelian and the Copernican, watching a sunrise, *see different things*. This is, of course, nonsense. What might be true is that the Aristotelian says of the sunrise 'I see that the sun is still orbiting the earth', while the Copernican says 'I see that the earth is still rotating on its axis'. The profundity 'The limits of my language are the limits of my world' is false. What is true is the triviality that the limits of my language limit what I can say of the world.

4. Permit me a true story. I was once told in all seriousness that when the concept 'person with an IQ two standard deviations above the mean' was invented, *new entities* were brought into being. It turns out (I replied) that there are two ways of making babies, the way we all know and love, namely love, and this new way, psychological theorizing! I was told that the new entities are not babies, indeed, are not persons with IQ's two standard deviations above the mean. I could gain no clear idea *what* they were.

■ | References

- Eddington, A. S. (1928) *The Nature of the Physical World*. (Cambridge: Cambridge University Press).
- Field, H. (1982) 'Realism and Anti-Realism About Mathematics', *Philosophical Topics*, 13: 45–69.
- Fine, A. (1984a) 'The Natural Ontological Attitude', in Lepin (1984): 83–107. [Reprinted in Fine (1986b), ch. 7.]
- Fine, A. (1984b) 'And Not Anti-Realism Either', *Noûs*, 18: 51–65. [Reprinted in Fine (1986b), ch. 8.]
- Fine, A. (1986a) 'Unnatural Attitudes: Realist and Instrumentalist Attachments to Science', *Mind*, 95: 149–79.
- Fine, A. (1986b) *The Shaky Game: Einstein, Realism, and the Quantum Theory* (Chicago: University of Chicago Press).

- Leplin, J. (ed.) (1984) *Scientific Realism* (Berkeley: University of California Press).
- Levin M. (1984) 'What Kind of Explanation Is Truth?' in Leplin (1984): 124–39.
- Pais, A. (1982) *Subtle is The Lord . . . : The Science and the Life of Albert Einstein*. (Oxford: Oxford University Press).
- Tarski, A. (1944) 'The Semantic Conception of Truth', *Philosophy and Phenomenological Research*, IV, 341–75.

9.1 Logical Empiricism	1227
9.2 Maxwell on the Ontological Status of Theoretical Entities	1229
<i>The Observational-Theoretical Distinction</i>	1229
9.3 Van Fraassen's Constructive Empiricism	1232
<i>Scientific Realism</i>	1232
<i>Constructive Empiricism</i>	1233
9.4 Van Fraassen's Reply to Arguments for Scientific Realism	1234
<i>Against Maxwell on the Observational-Theoretical Distinction</i>	1234
<i>On Inference to the Best Explanation</i>	1236
<i>Against Smart on the Success of Explanation</i>	1237
<i>Against Sellars's Thought Experiment</i>	1238
<i>Against Putnam's Ultimate Argument</i>	1240
9.5 Evaluating Constructive Empiricism: Musgrave and Others	1241
<i>Empirical Adequacy Once More</i>	1241
<i>The Observational-Theoretical Distinction Once More</i>	1243
<i>Realism and Explanation Once More</i>	1245
9.6 Laudan against Convergent Realism: Reference and Truth	1246
<i>Reference and the Success of Science</i>	1248
<i>Approximate Truth and Success</i>	1250
<i>Realism and Begging the Question</i>	1252
9.7 Brown on Explaining the Success of Science	1253
<i>Success, Realism, and Antirealism</i>	1253
<i>Verisimilitude and Success</i>	1256
9.8 Hacking's Experimental Realism	1261
9.9 Resnik's Evaluation of Experimental Realism	1263
9.10 Fine's Natural Ontological Attitude (NOA)	1267
<i>Against Realism</i>	1267
<i>The Nonrealist "Core": NOA</i>	1269
9.11 Evaluating NOA: Musgrave's Realist Reply to Fine	1272
9.12 Summary	1277

9.1 | Logical Empiricism

In the first half of the twentieth century, the logical positivists and their successors, the logical empiricists, approached the issue of scientific realism in a distinctive way by reflecting on the role of observation in two different types of scientific activity.¹ On the one hand, scientists seek to discover empirical laws, expressing generalizations about observable phenomena such as the motion of bodies or relations between the pressure and volume of a gas. On the other hand, scientists also formulate full-blown scientific theories that unify, explain, and predict such laws and their observational consequences in terms of forces acting on bodies, molecules of gases and their mean kinetic energies, and so on. Thus, it was natural for the logical empiricists to emphasize a distinction between the observational components of a theory, which refer to objects and properties that are directly observable, and the theoretical components, which apparently refer to objects and properties that are not directly observable.

The observational-theoretical distinction thus has two aspects, ontological and terminological (or linguistic). As an ontological distinction, it is used to mark off those objects, properties, and events that are directly perceivable from those that are not. For example, philosophers often distinguish between observable and unobservable objects. By contrast, the terminological distinction applies not to objects, properties, and events but to the language and vocabulary of scientific theories; it is usually made by calling those terms unique to a theory (such as *molecule*, *gene*, *electromagnetic field*, and so on) *theoretical terms*, and calling the other descriptive terms figuring in empirical laws *observational terms*.

The ontological distinction has been widely regarded by empiricist philosophers as having important implications for scientific realism. Ernst Mach, for example, (who was an important influence on logical positivism) argued that the aim of natural science is the economical description of experience. Since we can have no experience of unobservables, Mach concluded that they are at best convenient fictions introduced to simplify our theories. In the current idiom, Mach's view represents a kind of anti-realism about atoms, forces, and fields.

Switching now to the terminological distinction, consider an empiricist view of language, according to which the meaning of a term is exhausted by its connection with immediate experience. Given this view and the fact that theoretical terms are not connected directly with experience, the cognitive significance of theoretical terms would seem to be in jeopardy, they would seem to be meaningless. This sort of consideration

led philosophers of a broadly empiricist persuasion to accord theoretical terms a second-rate semantic status. The strongest version of such an account granted theoretical terms no essential role in scientific theory, regarding them as purely fictional terms to be eliminated altogether.² A somewhat weaker approach embraced by some logical empiricists treated theoretical terms as strictly speaking meaningless but nevertheless instrumentally useful in scientific theories; another approach, advocated by operationalists, insisted that theoretical terms so-called be given suitable content by explicit definition in observational terms. Finally, the weakest version of the empiricist approach granted theoretical terms a role in scientific theories but reckoned them to be only “partially meaningful” via their connection with observational terms. Thus, for example, Norman Campbell, Rudolf Carnap, and Carl Hempel viewed scientific theories as partially interpreted formal systems. On this account, the axioms of a theory from which observational consequences are deduced contain both observational and theoretical expressions, but only observational terms can be interpreted directly and fully; theoretical terms acquire a partial and indirect interpretation by contributing to the derivation of observational claims.

Logical empiricism may thus be viewed as a kind of antirealism: while the empirical claims of a theory in which observational terms figure are judged to be true because of the direct connection between those terms and real objects, events, and properties accessible to perceptual experience, a commitment to the meaningfulness of theoretical terms and the truth of purely theoretical claims in which they occur does not bring with it any commitment to their connection with unobservable objects, events, and properties.

Logical positivism is dead and logical empiricism is no longer an avowed school of philosophical thought. But despite our historical and philosophical distance from logical positivism and empiricism, their influence can still be felt. An important part of their legacy is the observational-theoretical distinction itself, which continues to play a central role in debates about scientific realism. In this commentary, we shall explore a number of questions about empiricism and scientific realism. Is the observational-theoretical distinction sustainable, and in particular, does it possess the ontological significance that many empiricists have taken it to possess? How exactly should we understand the role of theoretical terms in scientific theories? Does acceptance of a scientific theory bring with it commitment to the truth of theoretical claims? Can we explain the success of science without being realists about theories and the entities they postulate? Does the success of our scientific theories count as evidence for realism?

9.2 | Maxwell on the Ontological Status of Theoretical Entities

In his paper “The Ontological Status of Theoretical Entities,” Grover Maxwell expresses impatience with those inclining toward what we have called antirealism about theoretical terms and objects. His purpose is to isolate some influential strands of antirealism emerging from the logical empiricist tradition and to argue that all of them are implausible.

To begin, imagine (with Maxwell) a bit of fictional science in which, before the advent of microscopes, a scientist named Jones speculates about the mechanism by which diseases are transmitted from one person to another. Jones reasons—by analogy with observable mechanisms such as lice—that all infectious diseases are transmitted by some form of bug or other, most of which are unobservable. By dint of hard work, he comes to show that preventative measures aimed at killing such “crobes” could drastically lower the incidence of disease. But philosophically inclined scientists were worried: Jones had assumed the existence of objects that, according to his own theory, are unobservable. How should Jones’s theory be understood? Some claimed that the tiny organisms were mere fictions; others claimed that such a theory was at best an instrument, helpful to Jones’s thinking but not literally asserting the existence of unobservable entities. Still others claimed that sentences in which Jones’s theoretical terms appeared were permissible only to the extent that they could be translated into sentences containing just observational terms. But the day arrives when the microscope is invented, and Jones’s crobes are observed in great detail, different crobes being identified as the cause of different diseases. The antirealist philosophers respond in various ways. While many convert to realism, others stay their original course. Of these, (a) some espouse idealism or phenomenalism, insisting that all claims about physical objects, of whatever sort, be translatable into claims about immediately perceivable sense data;³ (b) others take the less drastic course of claiming that Jones’s crobes never had been unobservable in principle, since the theory did not entail the impossibility of finding a means of observing them; (c) still others insist that the crobes have not been observed at all: what is seen by means of the microscope is not a physical object or organism, but something far less substantial.

THE OBSERVATIONAL-THEORETICAL DISTINCTION

Maxwell argues against (a), (b), and (c), as they are manifested in actual positions philosophers have defended. In connection with (c), Maxwell cites a passage from the logical empiricist, Gustav Bergmann, who describes a certain empiricist rejection of atoms as entailing that strictly speaking “. . . even stars and microscopic objects are not physical things

in a literal sense," for when we look through our apparatus all we see "is a patch of color which creeps through the field like a shadow over the wall" (1055). Maxwell notes that if this line is taken, we cannot properly be said to observe garden-variety-sized physical objects through eyeglasses or windowpanes. Shall we say that we can only infer that it is raining, unless we raise the window and observe the rain "directly"? Consider, too, that modern chemistry tells us that there is a continuous transition from very small molecules (of water, say), through medium-sized ones (like proteins or nucleotides), to large ones (like diamonds or crystals of salt). The last of these are directly observable objects, but for all that they are no less genuine, single molecules. The lesson Maxwell encourages us to draw is twofold. First, we have no obvious criteria by which to draw a nonarbitrary line between the observational and the theoretical; while such line drawing as we can make will often prove convenient, its position will vary widely from context to context. Second, this continuous transition between observable and theoretical shows that the distinction has no ontological significance whatsoever. For surely an entity does not have real existence in one context and lack it in another, and surely objects seen through eyeglasses are not to be judged less real than (or to have a degree of existence inferior to that enjoyed by) objects observed directly by unaided vision.

Although he does not explicitly present them as such, Maxwell's reflections on (c) might be seen to recommend the following sort of argument: Since it is agreed on all hands that (1) there are observable objects about which we must be realists, and (2) given the continuous transition from observable to theoretical shows that there is no nonarbitrary line to be drawn between them, it follows that (3) we are not warranted in claiming that there is a class of in-principle unobservable objects about which we may be antirealists. Maxwell's claim that the observational-theoretical distinction "has no ontological significance whatever" (1062) will on this reading be understood as asserting that, whatever convenience the distinction may offer, it is not a distinction between things we can judge to be real and things we cannot.

Consider again view (b), according to which Jones's crobes never were unobservable in principle: on this view, only putative entities that are in principle impossible to observe must be rejected. Maxwell challenges this antirealist reliance on the notion of unobservability in principle. The current status of electrons is conceivably similar in many ways to Jones's crobes. For suppose we discover new entities able to interact with electrons in such a way that, with enhancements to our vision and by means of these entities, we are able to perceive various properties of electrons. However improbable such a scenario may be, it does not involve any logical or conceptual absurdity. Maxwell's point is that there are no a priori, purely philosophical criteria for separating the observable from the unob-

servable; to this extent, any descriptive term is a possible candidate for an observational term.

Might we improve our ability to draw a serviceable distinction here by focusing—as earlier philosophers have done—not on *observable* but on *observed* instead? On such a proposal, observational terms are taken to be those referring to what has in fact been observed. Construed narrowly, this would yield a different observational language for each language user: *sand* would be a theoretical term for some Eskimos, *snow* for some Floridians. Clearly this is too strong. But suppose we construe it broadly, as claiming that observational terms must be members of a kind, some of whose members or their properties have been observed. This is at once too weak and too strong: too weak because the similarity relations needed for specifying a kind are ubiquitous and easy to find; too strong because for any entity we can always specify a kind of which it is the only member. Maxwell concludes that such a strategy can only result in a retreat to some form of phenomenalism, which is position (a).

According to the phenomenalist version of antirealism (a), all claims about physical objects must be translatable into claims about immediately perceivable sense data. As Maxwell notes, scarcely any contemporary philosopher defends such a view. This is not to underestimate the relevance of sensory impressions themselves, but the fact remains that most statements of our common linguistic framework refer to public entities, not private inner states or objects. Indeed, this fact highlights the importance of an observational base in science as the ultimate grounds of confirmation. In scientific discourse the basic unit is not the observational term, but what Maxwell calls the quickly decidable sentence, a (nonanalytic) sentence that reliable, competent language users can quickly decide whether to assert or deny when reporting on some situation. How is it that we can and sometimes do quickly decide the truth or falsity of some observation sentence—a sentence whose only descriptive terms are those that may figure in a quickly decidable sentence? Maxwell argues that this is itself a scientific-theoretical question, the answer to which might well regard sense data as postulates of a theory. As before, the point is that such issues are not purely logical or conceptual. And thus, as before, it emerges that drawing the observational-theoretical line at any particular point is a purely contingent feature of our physiology, our current state of knowledge, and the instruments we happen to have available at the time. In light of this, Maxwell concludes that the observational-theoretical distinction is entirely void of ontological significance.

9.3 | Van Fraassen's Constructive Empiricism

Maxwell's is but one reply to the antirealist sentiments of logical empiricism in defense of scientific realism. Of course, if Maxwell's arguments fail, realism might still be tenable; for there may be other arguments to support it. Similarly, it would be an error to suppose that if Maxwell's criticism succeeds, then all varieties of antirealism have been shown to be false; for there may be other forms of antirealism.

The main arguments that have been offered in defense of scientific realism arise from criticisms of logical empiricism. But Bas van Fraassen defends a view that is at odds with both logical empiricism and scientific realism. In "Arguments Concerning Scientific Realism," drawn from his book *The Scientific Image*, van Fraassen proposes to do two things. His first task is to articulate a distinctive form of antirealism, which he calls *constructive empiricism*. His second task is to reply to the main arguments for scientific realism (including Maxwell's), from the perspective of his own antirealism. However untenable logical empiricism might be, the plausibility of scientific realism must be judged against the remaining anti-realist alternatives.

SCIENTIFIC REALISM

In rejecting scientific realism, what exactly is van Fraassen denying about the nature of science and its theories? An intuitive characterization of scientific realism would run as follows: the account of the natural world offered by science is true, and the entities it postulates really exist. Van Fraassen believes that this statement is too naive as it stands, since it entails that the scientific realist believes either that today's theories are correct or that science will eventually offer theories true in all respects. Still, the naive characterization is helpful in isolating those virtues of the scientific enterprise most prized by realists. In particular, realists are committed to the view that scientists seek to offer theories that are true and that the claims of such theories are made true by how the external world is. To accept a theory is thus (at the very least) to believe that its claims are true. Here then is van Fraassen's formulation of scientific realism:

Science aims to give us, in its theories, a literally true story of what the world is like; and acceptance of a scientific theory involves the belief that it is true. (1066)

Notice several things about this statement of scientific realism. First, it is suitably weak in two respects: it commits the realist to the claim that science aims at telling a true story, not that the story science tells is true, and while highlighting truth as the chief virtue of theories, it does not

entail that other virtues (such as simplicity, predictive and explanatory power, and so on) are irrelevant to a theory's success. Second, van Fraassen's reference to the goal of literal truth rules out the views of instrumentalists and other logical empiricists who assert that scientific theories may be taken loosely speaking as true if understood in the proper way but are nevertheless strictly speaking (i.e., taken literally) false or meaningless. Finally, acceptance of a theory requires only belief in its truth, without implying that scientists must be fully justified in such beliefs. Since justification comes in degrees, the realist can judge that a scientist's acceptance of a theory is strong or weak or somewhere in between.

CONSTRUCTIVE EMPIRICISM

Given this brief account of scientific realism, we can develop a general account of antirealism, with a view towards understanding van Fraassen's own distinctive version of it. According to the antirealist, the aim of science is not to offer literally true theories, and to accept a theory is not to believe it is true. Thus, while the realist will insist that someone who accepts a scientific theory is asserting its literal truth, the antirealist insists that to accept a theory is to do something less than that; it is to claim for it some virtue other than literal truth. Here antirealists will divide. As we have seen in the case of many in the logical empiricist tradition, an antirealist might claim that science aims at theories that are good because they are true when construed nonliterally.⁴ Alternatively, the antirealist might claim that theories are to be construed literally but that they need not be true to be good. Van Fraassen's antirealism is of this latter sort. On his literalist view, if a theory's statements include or entail "There are molecules," for example, then that theory says that there are molecules.

But how are we to understand the notion that one can take the language of science literally but deny scientific realism? Van Fraassen urges us to see that the literal construal of scientific language concerns our face-value interpretation of its meaning, while espousal of scientific realism concerns the extent of our belief. Thus, we are encouraged to see that taking theories literally is not believing they are true or believing that the entities they postulate are real. Instead, we are to adopt what he calls *constructive empiricism*:

Science aims to give us theories which are empirically adequate; an acceptance of a theory involves as belief only that it is empirically adequate. (1069)

A theory is empirically adequate when it "saves the phenomena"—when what it says about *observable* objects, events, and properties is true. The respect in which van Fraassen's antirealism departs both from logical empiricism and from scientific realism is thus apparent. To accept (hold) a theory is to claim that it accurately describes observable phenomena;

this does not entail that talk of theoretical entities is meaningless, nor does it entail that such entities are fictional or real. By distinguishing in this way between accepting a theory and believing it to be true, the constructive empiricist recommends a position of agnosticism about the theoretical. As van Fraassen says at the end of his book: "To be an empiricist is to withhold belief in anything that goes beyond the actual, observable phenomena. . . . To develop an empiricist account of science is to depict it as involving a search for truth only about the empirical world, about what is actual and observable."⁵

Van Fraassen emphasizes that, in addition to the epistemic dimension of constructive empiricism—concerning how far one's beliefs extend in accepting a theory—there is a pragmatic dimension as well. Accepting a theory *T* brings with it a commitment to explaining future phenomena in terms of the conceptual resources of *T*.

9.4 | Van Fraassen's Reply to Arguments for Scientific Realism

We noted earlier that the most influential arguments for scientific realism have come from critics of logical empiricism. Clearly, their success in establishing scientific realism cannot be judged simply by how well they dispense with the antirealism of the logical empiricists, if other forms of antirealism yet remain. Thus, van Fraassen's second task is to evaluate the main arguments for scientific realism in the light of his own constructive empiricism.

AGAINST MAXWELL ON THE OBSERVATIONAL-THEORETICAL DISTINCTION

How strongly do Maxwell's arguments support scientific realism? After distinguishing between terminological and ontological strains of the observational-theoretical dichotomy, van Fraassen poses two questions: Can we divide our language into a theoretical and nontheoretical part? Can we classify objects, events, and properties as being either observable or unobservable? To the first of these, recall that Maxwell answers "No." In particular, Maxwell claims that what counts as the fundamental unit of our observational base is not the term but the quickly decidable sentence and that an account of quickly decidable sentences will itself be a scientific-theoretical question. To this extent, then, even experimental and sensory reports threaten to involve theoretical elements. Van Fraassen agrees that, in the end, all language is to some degree theory-infected, but he denies that this shows anything about scientific realism. It may well be true that our language is influenced by theories, but this does not reveal

the extent of our beliefs about those theories. An extreme example best makes the point: to speak of the sun rising scarcely reflects a belief in the truth of Ptolemaic astronomy.

The more fundamental question is the second, concerning the possibility of dividing observable objects, events, and properties from unobservable ones. Maxwell answers this question negatively as well. Can van Fraassen follow Maxwell here? Presumably not. For in claiming that acceptance of a theory *T* involves believing only that *T* is empirically adequate, the constructive empiricist presumes we can distinguish between those statements of *T* that are about observable objects, events, and properties and those that are not; and this entails the intelligibility of judging whether or not they are observable.

Maxwell offers three arguments for realism: two against the possibility of distinguishing observable from unobservable objects and one against attributing any ontological significance to such distinctions as typically drawn. The first argument of the former sort leans heavily on the claim that we have no criteria by which to draw a nonarbitrary line between the observable and the unobservable, since there is a continuous transition from looking through air (with unaided vision), through a window, through binoculars, through a microscope, and so on. As we reconstructed the argument earlier (1230):

- 1 There are observable objects about which we must be realists.
- 2 No nonarbitrary line can be drawn between observable and unobservable objects.
- 3 Therefore, we are not warranted in claiming that there is a class of unobservable objects about which we may be antirealists.

Van Fraassen is prepared to grant Maxwell's premises, but he denies that its conclusion follows from them. In particular, premise (2) shows at most that *observable* is a vague predicate from which, even in conjunction with (1), nothing about realism follows. Despite the many sophisms that vague predicates have been used to generate, we are safe in using them as long as there are clear cases where they do apply and clear cases where they do not.⁶ At best, then, Maxwell is challenging us to find clear cases of unobservable objects. Van Fraassen claims that there are such clear cases. The purported "observation" of microparticles in a cloud chamber is one example. Here, a charged particle traverses the chamber to produce a visible trail of condensed vapor. In this case, the particle is indeed detected, and the detection is based on observation, but we can scarcely insist that the particle itself is being observed. Thus, even if we must be realists about observable objects, and even if the observable-unobservable distinction is vague, it does not follow that we are never warranted in claiming that there are unobservables about which we may be antirealists.

Maxwell's second argument against the observable-unobservable distinction is directed against the antirealist's use of *unobservable* as meaning "whatever the relevant theory entails cannot be observed." Since there are no conceptual or a priori grounds for supposing that there is any object that must remain unobservable regardless of the possible circumstances, the term *unobservable* marks out no class of objects for the antirealist at all. Van Fraassen claims that this argument is a mere distraction and changes the subject. Humans are measuring instruments subject to limitations that a final physics and biology will describe: it is to such limitations that the *able* in *observable* and *unobservable* apply. Just as we should not judge the Empire State Building to be portable because some giant might move it, so it is beside the point to note the logical possibility of detection devices outstripping the natural limitations of humans.

Finally, van Fraassen takes up Maxwell's argument against attributing any ontological significance to whatever of the observable-unobservable distinction might remain. The issue for the realist is the reality of entities postulated by science, and as Maxwell notes, something can exist even though it is unobservable. But van Fraassen emphasizes that even if this is so, it does not follow that the observable-unobservable distinction has no significance for realism. Specifically, scientific realism concerns the epistemic stance we take with respect to the claims of science, and on this issue, what we can observe is eminently relevant to what we should believe. The constructive empiricist claims that to accept a theory is to believe only that it is empirically adequate. And a theory's empirical adequacy depends entirely on what it says about observables. Thus, what a theory says about what is observable is relevant to deciding the case between realism and this form of antirealism.

Van Fraassen responds to four more arguments for scientific realism. In each case, van Fraassen's criticisms involve his constructive empiricist version of antirealism.

ON INFERENCE TO THE BEST EXPLANATION

Some have argued that the canons of rational inference require a commitment to scientific realism. In particular, we encounter in science itself a ubiquitous appeal to a rule of inference called *inference to the best explanation*. Given some evidence *E* and alternative hypotheses *H* and *K*, the rule says that we should infer *H* rather than *K* when *H* is a better explanation of *E* than is *K*. The argument under consideration is that following this rule consistently leads to scientific realism: to have good explanatory grounds for believing a claim *H* is to have good grounds for believing that the entities postulated by *H* exist. (Van Fraassen's homespun example: from the disappearance of my cheese, the midnight scratchings on the wall, and patter of tiny feet, I infer that there really is a mouse in the wainscoting.)

Does our widespread use of this pattern of inference commit us to believing in unobservable entities? Van Fraassen argues that it does not. Consider what it means to claim that we follow a certain pattern or rule of inference. It cannot mean that we explicitly formulate the rule to ourselves and consciously follow it. That is too strong: most people follow various rules of inference implicitly. Nor can it mean that we merely act in accordance with a rule (without conscious deliberation) in the sense that our conclusions could be reached by following that rule. That is too weak: any conclusion whatever could be reached by following the permissive rule that any conclusion can be inferred from any premise. Rather, the appropriate sense of following a rule must entail our willingness to believe what follows from that rule and our unwillingness to believe anything that conflicts with what follows from it. Thus, the claim that we follow a certain rule is an empirical hypothesis about what we are willing and unwilling to believe. The realist's claim that we always follow the rule of inference to the best explanation can be confronted with rival hypotheses. One of these rival hypotheses is van Fraassen's alternative: we are always willing to believe that the theory that best explains the evidence is empirically adequate. This brings with it no commitment to scientific realism whatever. Moreover, it accounts for the many cases where scientists argue for the acceptance of a theory on the basis of its explanatory power.

AGAINST SMART ON THE SUCCESS OF EXPLANATION

There are two kinds of arguments for scientific realism based upon explanatory power as a criterion for theory choice. One is offered by J. J. C. Smart; the other, by Wilfrid Sellars, will be taken up presently.

Smart argues that only scientific realism can account for the distinction between correct and merely useful theories. The realist can explain the usefulness of a theory *T* by appeal to the correctness of *T* or, if *T* is not true but is empirically adequate, by appeal to its relation to some rival that is correct. For example, consider two astronomers: the first is a realist about the Copernican hypothesis (*C*) but an antirealist about an empirically adequate version of the Ptolemaic hypothesis (*P*); the second, a thoroughgoing antirealist, accepts both *C* and *P* as useful but does not believe that either theory is true. The first astronomer can explain the usefulness of *C* easily enough (it is true); he can also explain the instrumental usefulness of *P*, about which he is an antirealist, in terms of its predictions agreeing with those of the true theory *C*. But how can the second astronomer, the thoroughgoing antirealist, explain the usefulness of the theories he accepts without believing that either *C* or *P* (or any other theory) is true?

Van Fraassen replies that empirical adequacy seems equally able to ground explanations of a theory's usefulness. Why think that an antirealist

can offer less of an explanation of the usefulness of *P* than the realist—as long as the antirealist begins, not with the premise that *C* is true, but instead with the premise that *C* is empirically adequate and thus gives accurate descriptions of planetary motion as observed from the earth? The realist defender of Smart's argument will of course respond that this only postpones the crucial question one step: for what explains the descriptive accuracy of *C* and, in particular, what explains the fact that all predictions of planetary motions fit *C*? It would seem that the antirealist must claim that the observable regularities simply *do* fit *C*, as a matter of brute fact, there being no need to posit entities and processes behind the phenomena to explain *why* *C* is a good theory. Smart believes that unless the observable regularities can be explained in terms of some deeper (unobservable) structure, we are left with having to believe in lucky accidents and sheer coincidences on a grand scale. But van Fraassen argues that these explanatory demands are unreasonable. For any theory will eventually bottom out, having to posit some regularities as basic or primitive, without receiving any further explanation at a deeper level. It is only a realist prejudice to suppose that these regularities must concern entities and processes that are unobservable. Moreover, van Fraassen is unprepared to grant the legitimacy of equating coincidence with having no explanation. He points out that inclination to call, say, your meeting me at the market a coincidence reflects only an absence of any plan to meet there, not an absence of explanations for each of us being there.

AGAINST SELLARS'S THOUGHT EXPERIMENT

It is common to distinguish individual observable facts from empirical laws and empirical laws from theoretical hypotheses. The observable fact that this water is boiling might be said to be explained by the empirical law that all water boils at 100°C, and this law itself might be said to be explained by theoretical hypotheses about latent heat and the energies of bonding between molecules and so on. Wilfrid Sellars has suggested that this picture is oversimplified, for our theories do not really explain such empirical generalizations; at best, they explain why observable things obey these regularities to the extent that they do. Moreover, strictly speaking, there may be no genuine regularities of observable phenomena at all: at most pressures water does not boil at exactly 100°C. But Sellars argues that this lack of strict regularity does not infect the theoretical level or genuine explanation would not be possible. Incompleteness or inferiority in our empirical generalizations requires some genuine universal regularities at the theoretical level of unobservables.

Sellars offers the following thought experiment to illustrate the point: Suppose at some early stage of chemistry we found that samples of gold that are observationally indistinguishable dissolve in aqua regia at different rates.⁷ We postulate two distinct microstructures, *A* and *B*, for gold to

explain this seemingly unpredictable variation in rates of dissolution by claiming that our various samples are really mixtures of distinct substances (whose microstructures obey strict regularities of dissolution). Clearly, no observational regularities can explain our experimental results; the explanation we offer would be impossible without an appeal to unobservable microstructures and regularities governing them. Since it is the aim of science to explain, science requires belief in unobservable microstructures even if the explanations they make possible have no further observational consequences.

Van Fraassen poses three challenges to the argument based on this thought experiment. First, is it really true that our microstructural postulate has no new observational consequences? If it is indeed true that *A* and *B* have distinct dissolution rates x and $x + y$, then all gold samples should dissolve at a rate somewhere between those limits, and any value between x and $x + y$ can be observed. But none of this is implied by our original data. So, it is false that the microstructural postulate has no new testable consequences at the level of observables.

Second, van Fraassen denies that science must explain in every case, even when the explanation brings with it no gain in empirical predictions. He appeals to the rejection of hidden-variable explanations for statistical phenomena in quantum mechanics. In classical statistical mechanics, probabilistic laws describe and predict phenomena that cannot, in practice, be analyzed at the level of single, individual molecules. Nevertheless, the theory still requires that individual particles obey completely deterministic laws, with each particle having a precisely definable position and momentum at every instant; only our ignorance prevents us from following the trajectory of each particle. In quantum mechanics, however, the fundamental descriptions of all systems are irreducibly probabilistic—no particle has a precisely definable trajectory, and we can predict only probability distributions for position and momentum. If quantum mechanics is true, then more precise information cannot be had, even in principle, since according to quantum mechanics, particles simply do not have fully determinate values of position and momentum at all times. Thus, quantum mechanics is fundamentally inconsistent with classical statistical mechanics, and no attempt to reduce quantum mechanics to a classical theory by postulating hidden variables (of the fully determinate classical kind) can possibly succeed.⁸

Sellars, the realist, recognizes this and hence does not insist on explanation in cases where none can be given. But van Fraassen has a further, more subtle, point. He argues that quantum mechanics accepts a principle that prohibits the adoption of theoretical variables when it is impossible to distinguish among them by any experiment or observation. Without this further principle, logical consistency alone would not rule out the possibility of some kind of hidden-variable theory that would be observationally equivalent to quantum mechanics. In fact, such hidden-

variable theories have been introduced by physicists such as David Bohm. They are not classical theories, but they do yield predictions that are identical to those of quantum mechanics. Indeed, the main argument that Bohm and others make for their theories is that they, unlike quantum mechanics, *explain* quantum phenomena. (The issue of choosing among observationally equivalent theories is explored at length in chapter 3.)

Third, van Fraassen offers a methodological rationale, as opposed to the explanatory one that Sellars defends, for deploying a microstructural hypothesis in the development of our imaginary science about the dissolving of gold in aqua regia. Perhaps, similar hypotheses might be posulated in connection with other metals, with the eventual result that new observational regularities are predicted about the behavior of alloys and amalgams. In short, perhaps the goal of science and its theories is not to explain, as the realists would have it, but to develop “imaginative pictures which have a hope of suggesting new statements of observable regularities and of correcting old ones” (1081–82).

AGAINST PUTNAM’S ULTIMATE ARGUMENT

Hilary Putnam has offered an argument for scientific realism that van Fraassen dubs the “Ultimate Argument.” (Some have called it the “miracle” argument.) According to Putnam,⁹ realism is the only adequate explanation for the success of science; if we do not believe that the claims of our theories are true, that many of its terms in fact refer to unobservable objects, then we can only reckon that the success of our theories is miraculous.

Van Fraassen is prepared to grant that an explanation of the success of science must itself be scientifically adequate. If we demand that science explain regularities in the world, the regular success of our scientific predictions also requires an explanation. Realists contend that our scientific predictions are often successful because they are based on theories that correspond with the world; how could such theories fail to describe and predict successfully? As expected, van Fraassen’s explanation of scientific success is not a realist one, but it is recognizably scientific in its own way. He offers a Darwinian explanation for the success of science: just as many species of organisms have struggled for survival, so have many theories, and just as organisms that were not adapted to their environment have not survived, so theories that do not have true observational consequences are cast aside. Thus, the success of science is no more a miracle than the survival of any species: empirically inadequate theories die out.

9.5 | Evaluating Constructive Empiricism: Musgrave and Others

Since the appearance of *The Scientific Image* in 1980, considerable attention has been paid to van Fraassen's criticism of realism and to his constructive empiricist version of antirealism. In addition to many discussions in philosophical journals, a collection of papers entitled *Images of Science* (1985) was devoted to an evaluation of van Fraassen's position. In the article "Realism Versus Constructive Empiricism" drawn from this volume, Alan Musgrave takes up three central issues in van Fraassen's account—the connection between truth and empirical adequacy, the theoretical-observational distinction, and the role of explanation in science.

EMPIRICAL ADEQUACY ONCE MORE

For theories about observables, or for those parts of our theories relating only to observables, truth and empirical adequacy amount to the same thing. They diverge only in the case of unobservables, where an empirically adequate theory could be either true or false, given all we can be said to know from the empirical facts. So the constructive empiricist recommends that we accept such theories but not believe them. Accordingly, the realist's epistemic commitment (in believing that a theory about unobservables is true) looks considerably riskier than that of the constructive empiricist. Thus it is that skeptical arguments against realism, to the effect that we could never be rationally warranted in believing theories about unobservables, get their purchase. In this connection, Musgrave argues that skeptical worries apply equally to constructive empiricism. Since an empirically adequate theory must save not merely all actually observed phenomena but all phenomena, past, present, and future, judging that a theory is empirically adequate goes beyond what we can know at any given time. To this same extent, then, we can never be warranted in accepting any theory as empirically adequate.

Van Fraassen insists that the threats (or potential costs) confronting realism and constructive empiricism are different. They are lower, for the antirealist, and "since it is not an epistemological principle that one might as well hang for a sheep as for a lamb" (1090), the epistemically responsible attitude to take is simply acceptance of a theory, not belief in its truth. Musgrave wants to suggest otherwise: if the risks of detection and subsequent penalty for two criminal acts are the same, the sensible criminal will opt for the one yielding the greatest gains. Since the realist takes no greater risk than the antirealist of being proved wrong on empirical grounds, the realist might as well be hung for the realist sheep (believing

our theories to be true) as for the constructive-empiricist lamb (accepting our theories only as empirically adequate).

The details of van Fraassen's assessment of this issue are found in later parts of his book not included in our selection. But his position might be summarized as follows: While the *risks* of realism and constructive empiricism may be comparable, the *penalties* are higher for realism. In particular, the realist must pay the penalty of rejecting empiricism, since the realist must concede that empirical evidence is not always the final arbiter of theory choice. Suppose two incompatible theories say exactly the same things about all matters observational. While the constructive empiricist can accept both theories, proclaiming each to be empirically adequate, the realist cannot believe them both to be true (on pain of contradiction). Since evidence cannot guide the choice between them, the realist is forced to allow nonevidential considerations into the decision. Realists commonly respond to the problem of empirically equivalent theories by arguing that putatively empirically equivalent theories can only really be evaluated by extending them—by embedding them in wider theories, where their equivalence will disappear. (A further discussion of issues involved in evaluating theoretical claims that evidence leaves underdetermined can be found in chapter 3.)

Before moving on to Musgrave's discussion of the theoretical-observational distinction and the role of explanation in science, let us note two further points about van Fraassen's concept of empirical adequacy. First, recall again van Fraassen's case of the mouse in the wainscoting. The realist emphasizes that we infer the existence of the unobserved entity (a mouse, in the example) as the best explanation of the observed phenomena (disappearing cheese, scratching noises in the wall). Van Fraassen claims that the inference succeeds in this example because the mouse is an observable thing after all: thus, he writes, "all observable phenomena are as if there is a mouse in the house" and "there is a mouse in the house" are in the present case "totally equivalent" (1077). This reflects van Fraassen's view, noted above, that in the case of observables the realist and constructive-empiricist positions coincide. That is, the realist's successful inference in the mouse example is a limiting case of van Fraassen's own proposal, since for theories solely about observables, theoretical truth and empirical adequacy amount to the same thing. Jeff Foss disagrees. In "On Accepting van Fraassen's Image of Science,"¹⁰ Foss argues that because van Fraassen equates empirical adequacy with saving the phenomena, an empirically adequate theory about observables might be false nonetheless. For example, it might be the case that "all observable phenomena are as if there is a mouse in the house" (1077) even though there is no mouse producing them. (Perhaps some other creature is devouring the cheese and scampering in the walls, or, just possibly, phenomenalism is true and there are no material objects at all.) Thus, when van Fraassen talks about "observable phenomena" and "saving the phenomena" he

must mean by *phenomena* something other than *observations* or *phenomenal appearances*. To avoid inconsistency in his characterization of empirical adequacy, van Fraassen must be using the term *phenomena* rather idiosyncratically to mean *truths about observable things*. But this leads to further trouble, for, from the realist's perspective, there are many truths about observable things that involve entities, processes, and properties that are not observable in van Fraassen's sense. For example, mice have bacteria in their gut, electrons in their tails, and electromagnetic fields emanating from their brains. Thus, realists would reject van Fraassen's claim that realism and constructive empiricism coincide when applied to theories about observable things.

Second, could we not undermine van Fraassen's commitment to constructive empiricism by applying constructive empiricism to itself? Recall that van Fraassen agrees that scientific realism can make sense of scientific activity and so, in that sense, is empirically adequate. His arguments aim to show, not that realism is false or inadequate, but that some other account of science, constructive empiricism, is also empirically adequate. But as John O'Leary-Hawthorne points out in "What Does van Fraassen's Critique of Scientific Realism Show?"¹¹ believing in constructive empiricism—as presumably van Fraassen does—is a far cry from merely proclaiming its empirical adequacy. According to constructive empiricism, we should believe only those claims whose truth or falsity we can settle by observation. Since the observable data cannot settle the truth or falsity of constructive empiricism, the consistent constructive empiricist should be an agnostic about his own doctrine. Thus, a commitment to constructive empiricism would seem to preclude van Fraassen from believing that constructive empiricism is true.

THE OBSERVATIONAL-THEORETICAL DISTINCTION ONCE MORE

The empiricist antirealist needs to distinguish between theory and observation. Van Fraassen agrees with the scientific realist that the observational-theoretical distinction cannot be defended, if the proposal is that scientific terms can be divided into two distinct classes. But van Fraassen does not concede that the continuous transition from direct observation (through the air) to indirect detection (through glass, binoculars, a microscope) on its own counts against the distinction between kinds of objects. The vagueness of the predicate *observable* does not entail the absence of a distinction between clear cases where it applies and clear cases where it does not. Thus, although looking at the moons of Jupiter through a telescope seems to van Fraassen to be a clear case of observation, the purported observation of microparticles in a cloud chamber seems to him clearly different: the microparticles are detected but are not themselves observed. And what if we had electron-microscopic eyes, allowing us to

observe what now we can only detect? Van Fraassen, recall, says that this changes the subject: the limitations on what is observable do not extend to the logically possible but only to what a final physics and biology would tell us are the inherent limitations on human organisms.

Musgrave argues that this reply is unsatisfactory. Physics and biology tell us that what is observable by humans varies from person to person and is a function of our peculiar evolutionary history. Such a person-relative and species-specific distinction can scarcely be granted the deep philosophical significance the antirealist needs to distinguish observables from unobservables. Van Fraassen claims that the observable-unobservable distinction itself enjoys no immediate ontological significance, but does have an epistemological significance. As his mouse-in-the-wainscoting discussion is meant to show, epistemic proprieties require that humans believe to be true only what a theory says about what can be observed. But Musgrave doubts that a person-relative, species-specific distinction of no ontological significance can bear this epistemological weight. He contends that the constructive-empiricist position has two elements: a methodological prescription (about what scientists ought to infer) and an empirical claim (about what scientists actually do infer) and that both are too strong. The methodological prescription is too strong, for given any plausible theory of evidential support, it is at least possible that there is better evidence for an explanation in terms of unobservables than there is for one in terms of observables. How reasonable is it to accept as evidence phenomena pointing to the existence of mice but not phenomena pointing to the existence of electrons on the grounds that we may one day see the mouse but not the electron? The empirical claim—that scientists in fact infer only the empirical adequacy of a theory, never its truth—is too strong as well: van Fraassen is prepared to agree that we can detect electrons but not observe them, yet how reasonable is it to suppose that a scientist who affirms that she has detected an object does not believe that the object exists?

Musgrave argues further that, worse than resting on unreasonable assumptions about scientific inference, van Fraassen's observable-unobservable distinction cannot be coherently made out. For suppose we grant his claim that our final physics and biology will tell us what is observable and what is not: a theory or set of theories *T* will say, among other things, that *A* is observable by humans and *B* is not. Shall we accept this verdict of *T*? We cannot accept it as true, since constructive empiricism requires that we accept as true only what *T* tells us about the observable, and surely "*B* is unobservable by humans" is not a statement about what is observable by humans. The consistent constructive empiricist cannot, then, believe that anything is unobservable by humans and hence cannot draw a workable distinction between the observable and the unobservable at all. (The constructive empiricist cannot retreat to a position according to which "observable by humans" is an observational pred-

icate which humans can, by observation, see to apply or not; for the construal of the observational-theoretical distinction as a claim about terms has been judged indefensible.)

REALISM AND EXPLANATION ONCE MORE

The realist claims not only that science explains facts about the world, but also that scientific realism itself explains facts about science. In particular, the realist believes that only scientific realism can explain the predictive success of science: without the belief that theoretical statements entailing the existence of unobservables are true and that such unobservables indeed exist, we can only judge the success of science to be a miracle. Musgrave reminds us of van Fraassen's reply to the Ultimate Argument: just as we do not ask why mice run from cats (since animals not fleeing their natural enemies are soon killed), so we do not ask why science is successful (theories that are not empirically adequate are soon cast aside).

Does this reply avoid the Ultimate Argument? Musgrave argues that it does not. On the one hand, the respectable scientist *will* of course ask the respectable question, Why does the mouse run from the cat? The answer is that the mouse perceives the cat, perceives it as an enemy, and runs. There is nothing un-Darwinian about such an explanation: it need not gloss the mouse's perception of the cat as its enemy in terms of intentional states in the mind of the mouse. Indeed the explanation can be readily wedded to what, on the other hand, is a more central Darwinian question: Why did that sort of mouse behavior evolve? The answer is as van Fraassen says—in an environment of mouse-hungry cats, it is cat-fleeing mice that tend to survive and thus perpetuate their cat-fleeing behavior. This explanation is not offered in place of the previous answer, but rather tells a story according to which the facts adumbrated in the previous answer are the result of the selective forces of nature. Musgrave goes on to urge that the case of success in science is similar. It is one thing to explain why some particular theory is successful and another to explain why existing scientific theories in general are successful. The latter explanation is the one that van Fraassen gives, and the realist and antirealist can equally accept it: unsuccessful theories do not survive. But that is not to explain why any particular theory is successful, and to this question the antirealist has given no answer.

Musgrave is prepared to grant that the Ultimate Argument cannot be construed simply as the claim that there can be no antirealist explanation of a theory's success. If, for example, a theory is devised precisely to predict phenomenal regularities we know to obtain—as the deferents and epicycles of Ptolemaic astronomy were designed to yield precisely those retrograde motions and periodic eclipses we do observe—then it is no miracle that such a theory should be successful in predicting what it does. But the case is different when a theory *T* designed to accommodate some range of

phenomenal regularities turns out to predict new, as yet unobserved regularities. The realist explanation is near to hand: the objects postulated by *T* exist, and what *T* says about them is true. What the antirealist seems forced to say is that *T* was formulated for one purpose and turned out miraculously also to be well adapted for some other purpose. Thus, Musgrave emphasizes that the Ultimate Argument for realism is most convincing when applied to *novel* predictive success. (For further discussion of the significance of novel prediction, see chapter 4.)

9.6 | Laudan against Convergent Realism: Reference and Truth

Our discussion thus far reveals the extent to which philosophers of science are prepared to evaluate realism and its rivals using criteria from the empirical sciences themselves. Scientists' preference for good explanations is chief among such criteria, and the Ultimate Argument represents its most general application by scientific realists: the undeniable success of science remains inexplicable unless the theories of science enjoy all or most of the virtues attributed to them by realists. These virtues, as we have seen, include the truth of scientific theories and the successful reference of their theoretical terms. Other virtues are frequently cited by realists. Given some commitment to the notion that science and its theories undergo not just change but genuine progress, realists will naturally want later theories to preserve and explain whatever approximate truth is captured by their predecessors. Science, it is said, increasingly converges on the truth about the natural world. This view has been called *convergent realism*.

In "A Confutation of Convergent Realism," Larry Laudan questions whether the connections between truth, reference, and success are as sound as the convergent realist believes them to be. To begin, he articulates five theses that make up what he calls "convergent epistemological realism" (CER):¹²

- R1 Scientific theories (at least in the 'mature' sciences) are typically approximately true and more recent theories are closer to the truth than older theories in the same domain.
- R2 The observational and theoretical terms within the theories of a mature science genuinely refer. . . .
- R3 Successive theories in any mature science will be such that they 'preserve' the theoretical relations and the apparent referents of earlier theories. . . .
- R4 Acceptable new theories do and should explain why their predecessors were successful insofar as they were successful. . . .

R5 Theses (R1)–(R4) . . . constitute the best, if not the only, explanation for the success of science. (1115–16)

Thesis (R5) expresses in general form the abductive (inference-to-the-best-explanation) argument for scientific realism. That is, realism explains the success of science, and so the success of science empirically confirms scientific realism. Laudan divides these abductive arguments into two general sorts. Arguments of sort I are based on the putative connection between the success of science and the likelihood that theories are true and genuinely referential. They take something like the following form:

- 1 If scientific theories are approximately true, they will typically be empirically successful.
 - 2 If the central terms in scientific theories genuinely refer, those theories will generally be empirically successful.
 - 3 Scientific theories are empirically successful.
-
- 4 (Probably) Theories are approximately true and their terms genuinely refer. (1116)

Abductive arguments for realism of sort II are based on the putative limiting-case relations between earlier and later theories and the likelihood that earlier theories are approximately true and genuinely referential. They run roughly as follows:

- 1 If the earlier theories in a 'mature' science are approximately true and if the central terms of those theories genuinely refer, then later more successful theories in the same science will preserve the earlier theories as limiting cases.
 - 2 Scientists seek to preserve earlier theories as limiting cases and generally succeed.
-
- 3 (Probably) Earlier theories in a 'mature' science are approximately true and genuinely referential. (1116)

Laudan's criticism of arguments of sort II, addressing the realist's broadly retentionist view of scientific progress and theoretical change, uses case studies from the history of science to suggest that the following sorts of relations between some theory *T* and its successor *T'* do not generally hold: *T'* entails *T*; *T'* retains the true consequences of *T*; *T'* preserves *T* as a limiting case; *T'* explains why *T* succeeded to the extent that it did; and *T'* retains reference for the central terms of *T*. (This concern with intertheoretic relations is discussed at length in chapter 8 and its commentary, and since Laudan's misgivings about realism are well represented

by his criticism of arguments of sort I, we have omitted his discussion of sort II arguments from the latter third of the paper included in the readings of this chapter.)

Let us focus on arguments of sort I. Laudan claims that its main premises (1 and 2 on p. 1116) cannot be sustained. His account may be divided into two parts: the first concerns the relation between reference and success; the second, the relation of approximate truth to success. We examine them in turn and offer an occasional realist response.

REFERENCE AND THE SUCCESS OF SCIENCE

The convergent realist who relies on (R2)—the thesis that terms in our scientific theories genuinely refer—in explaining the success of science must subscribe to the following claims (1117):

- S1 The theories in the advanced or mature sciences are successful.
- S2 A theory whose central terms genuinely refer will be a successful theory.
- S3 If a theory is successful, we can reasonably infer that its central terms genuinely refer.
- S4 All the central terms in theories in the mature sciences do refer.

According to the realist, reference explains success—that is, (S2) and (S4) entail and thus explain (S1)—while success warrants a presumption of reference—(S1) and (S3) provide warrant for believing (S4). Laudan is prepared to grant these relations; the central question is thus whether (S1) through (S4) are true.

On most readings, the claim of (S1) that theories of mature sciences are successful can be safely judged to be true. What of (S2)—the claim that a theory whose terms refer will be successful? Laudan contends that there is ample evidence that it is false. The history of science is replete with unsuccessful theories about atoms and molecules and genes and continents and so on. Moreover, it is easy to see how this could be the case. A theory's terms genuinely refer when the theory cuts the world at its joints, when it postulates entities that actually exist, but doing so scarcely entails or even makes likely the truth of that theory's statements about such entities. Thus, the content of a genuinely referential theory might easily be massively weighted on the side of falsehood.

As a speculative and critical digression, consider how a realist might respond to this assessment of (S2). If we encounter a theory massively weighted on the side of falsehood, how likely are we to believe that it succeeds in carving the world at its joints? More specifically, how probable is it that *no* likelihood is conferred on a theory by its terms genuinely

referring? That genuine reference should place so little constraint on likelihood of truth seems to presume that referential success could be largely accidental, having little or nothing to do with our judgments about the nature of the referents.

Some theories of reference deny this presumption. They claim that scientists cannot successfully refer out of the blue, willy-nilly; rather, successful reference (to members of a natural kind, for example) is ultimately grounded in some robust, nonaccidental connection between the use of a scientific term and its referent—a causal relation, say. If the realist adopts this view, and if this view is wedded to the doctrine that the properties of objects are (or are uniquely determined by their contribution to) the causal powers of those objects, then the realist can explain how our ability to refer to objects derives from our knowledge of the properties of those objects. If what grounds the possibility of genuine reference is precisely what grounds the empirical scientist's judgments about the properties objects have, then it would seem unlikely that successful reference could be disconnected altogether from the scientist's judgments about the nature of the objects. We can refer because we are causally connected with the properties objects actually have. While such connections to an object's properties in no way guarantees that all our judgments about that object will be true, the realist will urge nevertheless that successful reference in scientific theories is unlikely to be accompanied by wholesale falsehood. At the very least, then, such considerations suggest that the antirealist cannot be so sanguine about the absence or weakness of connections between reference and likelihood of truth, regardless of considerations about the nature of reference.

What about (S3), the realist's claim that success creates a reasonable presumption of reference? Laudan argues that many historical theories have been successful in generating explanatory accounts of various experimental phenomena, without being genuinely referential. In particular, a whole family of ether theories in eighteenth- and nineteenth-century physics and chemistry were used to account for a wide range of phenomena—from the heat of reactions in chemistry to the refraction, interference, and polarization of light waves in optics. Judging from such historical cases, Laudan concludes that (S3) must be rejected. Nor will it do to weaken (S3), to claim that (only) *some* of the terms of a successful science can be inferred genuinely to refer. In the first place, doing so is not open to the typical realist, who claims not that evidence for a theory is evidence for only part of that theory (the parts about observable things, for example), but rather that evidence for a theory is evidence for everything it claims, for the truth of that theory as an account of everything it is about. If the tests to which we subject our theories test only parts of them, then even highly successful theories may have central terms (at the deepest theoretical level, perhaps) that do not refer and may make claims that, as yet untested, we would have little grounds for judging to be true. Moreover,

weakening (S3) in this way threatens the realist's rationale for holding (R3)—the thesis that theory succession is and should be retentive. The underlying motivation for (R3) is the realist's conviction that, because the earlier theory was successful, its terms must have been referential, and that retaining referential terms is a constraint on successor theories. If we weaken (S3), this constraint is undermined.

APPROXIMATE TRUTH AND SUCCESS

The other central element in arguments of sort I concerns the relation of truth to success. Realists typically claim that, while full theories (as, say, conjunctions of statements) may well be strictly false, they are nevertheless close to the truth or enjoy some form of verisimilitude. As Laudan puts it:

The claim generally amounts to this pair:

T1 if a theory is approximately true, then it will be explanatorily successful;
and

T2 if a theory is explanatorily successful, then it is probably approximately true. (1124)

In discussing the “downward path” (1123) from approximate truth to explanatory success in (T1), Laudan's primary misgiving concerns the notion of approximate truth. Realists have offered too little by way of an analysis of this concept to show that, as (T1) asserts, approximate truth entails explanatory success. Surely (T1) is not obviously true. Even if the approximate truth of a theory is taken to mean that the size of the set of true claims entailed by the theory is vastly larger than the set of its false entailments, it may yet be the case that most of the theory's consequences thus far tested are false. Without some coherent account of approximate truth that can be seen to entail explanatory success, we are within our epistemic rights to deny (T1).¹³

Laudan argues that the “upward path” (1126) from explanatory success to (probable) approximate truth expressed in (T2) fares no better. He thinks that the main problem, for the realist, is the assumed connection between genuine reference and approximate truth, for Laudan asserts that:

L “A realist would never want to say that a theory was approximately true if its central theoretical terms failed to refer” (1126).

If there were no genes, then genetic theory—however well confirmed it might be—would not be approximately true. According to (L), a necessary condition, especially for the scientific realist, for a theory being close to the truth is that its central explanatory terms genuinely refer. But the history of science is replete with examples of theories that were both suc-

cessful and nonreferential with respect to their central explanatory terms: among Laudan's examples are the crystalline spheres of ancient and medieval astronomy, the phlogiston of eighteenth-century chemistry, and the optical and electromagnetic ether of nineteenth-century physics. In all such cases, the theories were successful and well confirmed but contained central terms that we now believe not to refer.

In response to Laudan's contention that many successful theories failed to refer, the realist might emphasize again that (R1) through (R5) apply only to theories of a mature science—theories offered at some point sufficiently far into the life of a discipline that all the relevant background theories are reasonably entrenched and well confirmed. Many of the theories on Laudan's list, being drawn from immature sciences, would then fail as counterexamples to the connection that realists claim to hold between approximate truth and explanatory success. Laudan finds this maneuver unsatisfactory in two respects. First, it threatens to make (R3) and (R4) vacuously true. For if the mature sciences are *defined* as those in which correspondence and limiting-case relations hold between successive theories, then (R3) and (R4) become true by definition. Second, the restriction to mature sciences undermines the realist's aim to explain why science in general is successful. No such general explanation will have been given on the restricted account if indeed there are, as Laudan has argued, successful theories in immature sciences.

Before concluding our discussion of Laudan, it is worth noting that Clyde L. Hardin and Alexander Rosenberg see no reason realists should grant thesis (L), that "a realist would never want to say that a theory was approximately true if its central theoretical terms failed to refer."¹⁴ Laudan's own example of the gene and genetic theories serves as a useful case here. (See the section "Kitcher on Reduction, Classical Genetics, and Molecular Biology" in the commentary on chapter 8). Mendel's nineteenth-century theory is at the beginning of a sequence of theories enjoying increasing degrees of success, a sequence many biologists would claim is converging on the truth. Despite being credited with approximate truth, despite the ease with which it can be modified to generate increasingly accurate and complete genetic theories, Mendel's theory used the term *gene* to refer to entities that do not in fact exist.¹⁵ That is not to say that the theory was incorrect in ascribing to the fundamental genetic material most of the functions it did; rather, the functions thus ascribed to genes are now parceled out to different sequences and complex combinations of DNA. Indeed, in much of contemporary molecular genetics the use of the term *gene* has simply been dropped in favor of terms better suited to articulate the diverse units of hereditary phenomena and their functions. The causal role Mendel accorded to genes has been divided up among other entities, entities that function together to give the false impression that a unitary genetic item is responsible for all aspects of genetic transmission and expression. Thus, a realist can explain the plausibility of

Mendel's theory, and its approximate truth, by appeal to those very diverse functions of diverse objects that reveal that the term *gene* does not refer.

That is one way the realist might go. Another way is to construe Mendel's theory (perhaps with the aid of theories of plural reference or partial reference) as referring successfully after all to configurations of DNA and their polypeptide products, despite Mendel's ignorance of DNA and its role in protein synthesis. Here, unlike the previous account, the realist can opt to sever successful reference from the detailed beliefs of the scientist, granting the Mendelian theory its undeniable measure of success in the light of newer approximations to the truth.

The realist can either trace out the relations between Mendel's theoretical claims and those of modern molecular biology in a way that explains the approximate truth of the former despite their failure of reference or adopt a theory of reference able to preserve the referential success of the term *gene* as described. There being as yet no final and fixed theory of reference in philosophy, a good deal rests on one's views in this area. Nonetheless, whichever way one chooses to understand the relation of Mendelian to molecular genetics, one can be a realist about the success and approximate truth of the earlier theory.

REALISM AND BEGGING THE QUESTION

The realist strategy that Laudan has been attacking attempts to link empirical success with likelihood of approximate truth and reference. In his final remarks Laudan argues that, leaving aside the status of its premises, such arguments can—from the point of view of antirealists—only be seen as question begging. Antirealists deny that we are warranted in judging any theory true on the basis of its yielding true observational consequences. To suppose that we are is to commit the fallacy of affirming the consequent, since a false theory can have true consequences.¹⁶ Now, realists have been concerned, not with any scientific theory in particular, but with the truth of realism, with the truth of a claim about scientific theories generally. That is, realists have argued that we can reasonably judge scientific realism to be true because it yields true consequences. (If our scientific theories are approximately true and genuinely referential, then they will enjoy empirical success, and they do enjoy empirical success.) But to suppose that realism can be defended in this way is to beg the question against the antirealist, who already denies that the empirical success of any theory is evidence of its truth.

9.7 | Brown on Explaining the Success of Science

In “Explaining the Success of Science,” James Robert Brown sets out to reevaluate some of the more prominent ways in which philosophers of science have sought to explain why science is successful. Brown’s paper is especially helpful in two important respects. First, it serves as a useful review of our discussion of scientific realism thus far and in doing so sets the stage for Ian Hacking’s account of experimental realism (below). Second, as Laudan’s discussion emphasizes, the notions of *success* (as applied to theories) and *approximate truth* (or truthlikeness), while central to the debate among scientific realists and antirealists, are obscure and difficult to articulate. Brown offers a clearer statement of the notion of success than we have encountered thus far and introduces us to a more precise analysis (due to Newton-Smith) of verisimilitude and its relation to success.

SUCCESS, REALISM, AND ANTIREALISM

Suppose we want to resist the conclusion that the success of science is a miracle, a mystery incapable of being explained. To evaluate various approaches to explaining the success of science properly, we need to specify what we have in mind when we say that theories are successful. Brown suggests that the following features are implicit in the traditional view:

- 1 [Successful theories] are able to organize and unify a great variety of known phenomena. [empirical adequacy]
- 2 This ability to systematize the empirical data is more extensive now than it was for previous theories. [increasing adequacy]
- 3 A statistically significant number of novel predictions pan out; that is, our theories get more predictions right than mere guessing would allow. [novel predictions] (1136–37)

The main issues of debate concern not the success of science per se, but rather the broadly ontological consequences¹⁷ of various approaches to explaining its success. Appeals to truth and reference in explaining science are of course congenial to a realist about the ontological commitments of theories, while antirealists favor explanations they judge to be ontologically minimal or neutral. We have encountered important examples of both approaches.

One prominent realist explanation for the success of science that we have seen is based on the claim that theories are true or approximately true. Since the empirical success of science would follow from the truth of its theories, and since it is true that science is successful, then (since whatever is deduced from truths is true) the assumption that theories are indeed true explains their success. As Brown reconstructs it (a version of

what we earlier called the Ultimate Argument), for any theory *T* and any true empirical observation *O* following from *T*:

- 1 Conclusion *O* can be deduced from theory *T*.
- 2 *O* is observed to be the case.
- 3 If *T* is true then the argument for *O* is *sound* and so *O* *had* to be true.
- 4 If *T* is false then the argument for *O* is *merely valid* and the probability of the arbitrary consequence *O* being true is very small. (I.e., it would be a miracle if *O* were true.)

∴ The argument for *O* is probably sound.

∴ *T* is probably true. (That is, even *T*'s theoretical statements are probably true.) (1138)

A prominent antirealist explanation for the success of science that we have seen is the Darwinian account defended by van Fraassen: Just as there are many species struggling for existence, so too there are many theories competing for our acceptance; and just as species that fail to adapt to their environment become extinct, so too theories that fail to make true observational predictions are given up. So there is no need to assume that theories are true or approximately true in explaining the success of science.

Are either of these explanations right? (They cannot both be right, though of course both could be wrong.) Brown examines first the anti-realist Darwinian case. The relevant question is whether it accounts for all three properties of successful theories identified earlier. Brown has two misgivings. First, while van Fraassen's account appears to explain (1) and (2)—a successful theory organizes and unifies the phenomena and systematizes the data better than earlier theories did—it pretty clearly fails to explain (3)—why do our theories make correct novel predictions more often than one would expect on the basis of mere chance? The Darwinian analysis breaks down here: the analogue to novel predictions is a radical change in environment, which most species do not survive. Brown's second, more general misgiving is that the Darwinian account is implicitly committed to the notion that rational theory choice and empirical success go hand in hand. Having made this assumption, it is unsurprising that the theories we accept have properties (1) and (2), since according to the empiricist we choose our theories on that very basis. But, according to Brown, rational theory choice does not rest solely on empirical adequacy. Conceptual, metaphysical, and aesthetic concerns can also play a crucial role in deciding among theories; consequently, anyone except the most extreme empiricist (a logical empiricist, say) must grant the possibility that the most rationally acceptable theory might not be the one that is most successful empirically. In light of this result, we should be less quick to concede that van Fraassen's Darwinian account adequately answers the

questions, Why do the theories we accept get so much right empirically? and Why do new theories get more right than old ones?

The realist case is somewhat more difficult to assess because most realists tie the truth of theories, and the increasing approach to truth by successive theories, to the notions of reference and preserved reference (from earlier to later theories). Recall, for example, Putnam's claim that the realist sees genuine reference, the approximate truth of mature theories, and preserved reference "as part of the only scientific explanation of the success of science . . ." (1140). Leaving the issue of approximate truth aside for the moment, must we agree that the terms in successful mature theories typically refer?

Brown, like Laudan, thinks not. Consider the very simple theory Brown offers (1140):

T₁: JRB went to Dubrovnik in April.

Let us suppose that this is quite a successful theory and that all its terms refer. Is the fact that all its terms refer sufficient for its success? Surely not. Consider Brown's next theory, which is very unsuccessful:

T₂: JRB did *not* go to Dubrovnik in April.

Here then is a counterexample—a theory all of whose terms refer that is unsuccessful. Brown notes that less artificial counterexamples can be readily drawn from the history of science. The history of atomism is filled with unsuccessful theories, despite the fact that (we believe) the term *atom* refers. And while genuine reference is not sufficient for success, neither is it necessary: as Laudan suggests, phlogiston theories, caloric theories, ether theories, and many others all had their heyday, even though by our present lights we judge their terms not to refer.

One point in passing, encountered earlier: As in the case of the term *gene*, the realist is not committed to agreeing that *atom* referred in earlier, unsuccessful theories. One can endorse quite fine-grained constraints on the nature of successful reference without giving up the realist position that the terms of current theories refer or that whatever success was enjoyed by earlier theories was not tied to genuine reference. Suppose I claim that there are *F*s possessing the (observational or theoretical) properties *G*, *H*, *J*. . . . Your showing that I have failed to refer with the term *F* leaves untouched my claims that there is something that is *G*, that is *H*, that is *J*, and so on. For all my failure of reference, I may have said quite a few true things about the world. (Presumably this is the line taken by those realists who claim that Mendelian genetic theory was approximately true despite the term *gene*'s failure to refer.)

VERISIMILITUDE AND SUCCESS

While the succession of theories in a scientific discipline undoubtedly involves many respects in which later theories are better than earlier ones, it is approximation to truth, or truthlikeness (verisimilitude), that figures most centrally in realist accounts of progress. Brown aims to clarify this difficult notion by examining William Newton-Smith's analysis of it in some detail.¹⁸ Newton-Smith's analysis has two main virtues. First, it does not depend on reference being preserved across successive theories. Second, it makes explicit the connection between increasing verisimilitude and scientific progress. In particular, Newton-Smith proves that theories with greater verisimilitude (according to his definition of that concept) must have a higher probability of observational success.

Brown introduces several key notions and assumptions in Newton-Smith's account. First, a theory is defined as a set of claims, the fundamental postulates and hypotheses of the theory together with all their deductive consequences. Second, Newton-Smith assumes that the consequences of theories can be enumerated recursively or, in other words, that they can be generated by an algorithm in a numbered sequence. Suppose that the consequences of T_1 are enumerated as $t_1 = t_1^1, t_1^2, t_1^3, \dots$ and the consequences of T_2 are enumerated as $t_2 = t_2^1, t_2^2, t_2^3, \dots$. Newton-Smith stipulates that all logically true sentences have been deleted from these lists of consequences and that once a sentence appears in the list, all subsequent sentences that are logically equivalent to it are eliminated. Third, a theory is said to decide p if the theory (that is, its entire set of consequences) contains either p or not- p . Fourth, the relative content of a theory is defined comparatively: of two theories, T_1 and T_2 , the one with greater content is the one that decides more sentences.¹⁹ Finally, the notion of relative truth—roughly, the number of truths of one theory compared with those of another—is elucidated as follows: from t_1 (the list of consequences of T_1) and t_2 (the list of consequences of T_2), we construct an infinite sequence of ratios. The n th term of this sequence is the number of truths among the first n sentences of t_1 divided by the number of truths among the first n sentences of t_2 . So, for example, if 3 out of the first 10 sentences of t_1 are true, and 7 out of the first 10 sentences of t_2 are true, then the 10th term of the sequence is $3/7$. (If the ratio were greater than 1, we would invert it to ensure that each term in the sequence is a fraction between 0 and 1.) Of course, deciding which sentences in t_1 and t_2 are true is no easy matter, especially since many of the consequences of real scientific theories will be highly theoretical. So Newton-Smith proposes that the only practical way to judge their truth is from the perspective of some third theory T_3 , a plausible candidate for which would be a theory that we presently regard as true. This leads Newton-Smith to his definition of relative truth (which differs only in minor ways from the version given by Brown on page 1143):

T_2 has a greater truth relative to T_3 than T_1 if and only if the infinite sequence of ratios giving the ratio of truths in T_1 to the truths in T_2 judged by reference to T_3 , has a limit greater than $\frac{1}{2}$ which is unaffected by reasonable place-selection.²⁰

Then, using his definitions of relative truth and relative content, Newton-Smith defines verisimilitude as follows:

T_2 has greater verisimilitude than T_1 if and only if both: (1) the relative content of T_2 is equal to or greater than that of T_1 ; (2) T_2 has greater truth relative to T_3 than T_1 .²¹

Intuitively, a theory has greater verisimilitude than another if it says more about the world and more of what it says is true (at least, as judged from the standpoint of a third theory). Thus, verisimilitude is connected to observational success. If T_2 has greater verisimilitude than T_1 , then the chances of any arbitrary sentence of T_2 being true are greater than the chances of any arbitrary sentence of T_1 being true, and since the arbitrary sentences of a theory include its observational consequences, T_2 is likely to have more observational success.

Before proceeding, there is an error in the definition of greater relative truth that Newton-Smith has since corrected. Why should the ratio of truths of T_1 to T_2 have a limit greater than $\frac{1}{2}$ in order for T_2 to have greater truth than T_1 ? If, for example, T_1 and T_2 contained the same number of truths, the limit of truth ratios would be equal to 1 (and hence greater than $\frac{1}{2}$), but obviously neither theory would have greater truth than the other. When theories do differ markedly in truth content, the limit of truth ratios will be close to 0. Thus the definition of greater relative truth should require that the limit be less than 1, not greater than $\frac{1}{2}$.

Subject to the minor correction just noted, Newton-Smith's analysis of the approximate truth, or truthlikeness, of a theory has a number of attractions, including its intuitive appeal and relative simplicity. Despite these attractions, Brown offers three criticisms of Newton-Smith's analysis. First, asks Brown, how well does it account for the three features of successful theories—empirical adequacy, increasing adequacy, and novel predictions? The last two, the increased systematic and predictive power of later over earlier theories, seem to be explained. But, Brown points out, Newton-Smith's proposal says nothing about why present theories get so much right; indeed, his account is consistent with present theories doing quite poorly at organizing data and predicting truths.

Second, Brown thinks that Newton-Smith's definition of verisimilitude is too crude to do justice to scientific theories that assign numerical values to physical constants. Brown explains this criticism using a simple

example of two miniature theories T_1 (Theory 1) and T_2 (Theory 2), both of which are false (1144):

The Truth: John is 180 cm tall.

T_1 : John is 178 cm tall.

T_2 : John is 179 cm tall.

Intuitively, we want to say that theory T_2 is closer to the truth than theory T_1 . But, Brown argues, Newton-Smith must regard both theories as “totally and equally false” because he measures truthlikeness by judging individual sentences either true or false “and never something in between” (1144). Thus, Brown suggests that a theory of verisimilitude should assign degrees of truthlikeness to individual sentences. If we imagine, say, that successive theories differ only in their estimations of the value of a certain physical constant (say Planck’s constant or the gravitational constant), then surely the relevant difference among them is how close their particular claims about the value of the constant are to the truth.

The argument in Brown’s second objection has a fundamental flaw. It is not the case that Brown’s theories T_1 and T_2 are “totally false” simply because each consists of a single false sentence. Remember that Newton-Smith defines a theory as the entire set of its deductive consequences, and a false theory—even one stated in a single sentence—can have many true consequences. We can illustrate this point by enumerating some of the consequences of T_1 and T_2 . For the sake of simplicity, we shall restrict our attention to whole numbers, and write true sentences in regular type, false ones in *italics*.

T_1	T_2
John is taller than 1 cm.	John is taller than 1 cm.
John is taller than 2 cm.	John is taller than 2 cm.
John is taller than 177 cm.	John is taller than 177 cm.
<i>John is not taller than 178 cm.</i>	John is taller than 178 cm.
<i>John is not taller than 179 cm.</i>	<i>John is not taller than 179 cm.</i>
John is not taller than 180 cm.	John is not taller than 180 cm.
John is not taller than 181 cm.	John is not taller than 181 cm.

Even when restricted to whole numbers, these two enumerations are not complete. For example, each theory also entails the false sentence “John is shorter than 180 cm,” a sentence that, it should be noted, is not logically equivalent to any other single sentence in either list. But the main point remains unaffected by this incompleteness: although T_1 and T_2 are false, they are not “totally false” as Brown alleges.

Despite the flaw in Brown’s reasoning, we can still salvage his philosophical criticism of Newton-Smith’s proposal. For if we apply Newton-

Smith's definition of verisimilitude (incorporating the correction noted earlier) to the restricted lists of consequences just enumerated, it follows that the limit of the sequence of ratios giving the ratio of truths in T_1 to the truths in T_2 is 1. Thus, since the relative content of the two theories appears to be the same, Newton-Smith would judge that T_2 does not have greater verisimilitude than T_1 , which conflicts with our intuitive judgment that T_2 is closer to the truth.²²

Finally, Brown notes some historical cases in which later theories made progress by focusing on narrower domains than their predecessors had. For example, progress was made when theories that combined astrological and astronomical claims were replaced by astronomical theories that said nothing about astrology. On Newton-Smith's account, a later theory with less content than its predecessor cannot have greater verisimilitude than the theory it replaces, despite the fact that, intuitively, the later theory might be closer to the truth.

Up to this point we have been concerned with the concept of verisimilitude and the entailment from greater verisimilitude to the likelihood of greater observational success. It is now time to consider the argument in the other direction, from greater observational success to the likelihood of greater verisimilitude. Earlier, we encountered Laudan's antirealist objection that such abductive inferences beg the question. As Brown expresses the objection, antirealists are already unconvinced by hypothetico-deductive (H-D) inferences of the form (1145):

Theory \rightarrow Observation

Observation

\therefore (Probably) Theory

Thus, realists can only beg the question against antirealists by arguing:

Greater verisimilitude \rightarrow Greater observational success

Greater observational success

\therefore (Probably) Greater verisimilitude

Brown reasons that whatever flaw antirealists see in such realist arguments, it cannot arise simply from the problem of induction, for most antirealists accept inductive reasoning as legitimate. Van Fraassen, for example, while resisting inferences to the existence of unobservables, is nevertheless willing to infer that theories are empirically adequate. These judgments go beyond what is observed, since they entail the truth of all observational consequences. Indeed, to stop making inductive inferences would be to stop doing science altogether.²³ So the issue dividing realists and antirealists, in Brown's assessment, must be a deeper disagreement about where to draw the line among ampliative inferences. The hard-line

skeptic about induction stops at the evidence itself and refuses to go any further; van Fraassen allows the inference to the conclusion that a theory *T* is empirically adequate but withholds judgment about the existence of unobservable entities posited by *T*; others of a deflationary-realist bent (such as Ian Hacking, discussed below) allow that the unobservables posited by *T* exist yet resist the claim that *T* itself is true;²⁴ while full-blown scientific realists go on to commit themselves to the existence of unobservables and to the truth of *T* as well.

The disagreement about which ampliative inferences are legitimate makes it very difficult for realists to convince their opponents that scientific theories are true or close to the truth, or that later theories have greater verisimilitude than earlier ones. But traditional realists use these claims about truth and verisimilitude to infer, and thus explain, the success of science. In the final section of his paper, Brown speculates about how realists might avoid this problem. What is needed, he suggests, is a different style of explanation. Proponents of both the deductive-nomological and inductive-statistical models in the tradition of Hempel (see chapter 6) view explanations as arguments in which a conclusion is inferred from premises: given the premises (the explanans), the conclusion (the explanandum) then follows with certainty or with high probability. That is, a good explanation offers a sufficient or nearly sufficient condition for what is being explained. And here lies a weakness in the traditional realist position, for truth is not sufficient (or necessary) for the success of a scientific theory, nor does the truth of a theory make its success highly probable. (Remember that, in this context, success means not merely having true consequences but being able to organize and unify a great variety of known phenomena.) So the Hempelian requirements of deducibility and high probability are not met. Still, truth is relevant to success, just as Jones's syphilis is relevant to his paresis: in both cases, the explanandum is made more probable than it would have been otherwise.

The statistical-relevance view of explanation to which Brown appeals here bears some affinity to what is sometimes called the narrative style of explanation. Narratives explain events and facts by telling a story that makes them intelligible. Why does the giraffe have a long neck? The narrative approach would explain this by offering a story—in this case an evolutionary story according to which ancestors of the modern giraffe could survive only by feeding on trees. The story is not offered with any confidence as being true, but as a genuine possibility, a hypothesis that, if true, would make intelligible the phenomenon to be explained. Like statistical-relevance explanations, narrative explanations describe factors that are merely relevant to, not sufficient for, the explanandum; unlike statistical-relevance explanations, narrative explanations offer these factors simply as conjectures, not as established facts.²⁵

Brown's speculative proposal, then, is that realists do have an explanation for the success of science. Truth is the explanation, but the ex-

planatory style is narrative. We cannot claim to know that our theories are true; rather, their truth is conjectured, hypothesized. And even if our theories are true, their success does not automatically follow. But if our theories are true, then their truth will make a difference to their success. According to Brown, that is the most we can hope for in explaining the success of science. Moreover, we cannot presume that success (the explanandum) confirms realism (the explanans) because narrative explanations are too weak to generate any significant degree of confirmation. Thus, the success of science does not justify one in believing that realism is true. For that, Brown suggests, we need considerations of a quite different sort.

What sort of considerations might those be? One answer to this question is given by Ian Hacking's experimental realism, to which we now turn.

9.8 | Hacking's Experimental Realism

We mentioned in passing above a kind of deflationary realism that expresses a commitment to the existence of unobservable, or theoretical, entities without making any commitment to the truth of scientific theories. Ian Hacking is such a realist. In his paper "Experimentation and Scientific Realism," he claims that most discussions about scientific realism and antirealism, conducted as they are at the level of theories, are necessarily inconclusive. Realism about theories—about the truth of theories as claims about the natural world—is based only on an optimism attending our continued invention of successful new theories. Realism about entities, on the other hand, asserts the existence of (at least) the unobservable entities of physics, and Hacking claims that we can justify such assertions at the level of experimental practice. We are warranted in being realists about entities once we have manipulated them, measured them, and intervened in their causal processes. The relevant considerations for the deflationary realist thus involve not theorizing but experimenting.²⁶

One might wonder if the distinction between theory realism and entity realism can be made out. If one believes a theory, then one surely believes in the entities it postulates, and if one believes in certain entities, then one surely believes truths about them. Hacking resists this objection. He contends that the vast majority of experimental physicists are convinced of the existence of "inferred" and "unobservable" entities but that few believe their theories are literally true. Typically, experimenters will remain agnostic about the many possible (and mutually incompatible) theoretical frameworks in which claims about electrons, say, can be cast. This is not simply a sociological report about experimentalists. The point is rather that experimentalists can be antirealists about scientific theories

and models, viewing them—rather as instrumentalists might—as tools for thinking and calculating while at the same time believing in the existence of the entities they investigate. Their enterprise would be incoherent if they did not: realism about entities is a necessary condition for the coherence of scientific experimentation.

Before presenting an example of scientific experimentation to illustrate his case, Hacking makes two important points, about reference and about what he calls *interfering* or *intervening*. The first point (as we shall interpret it) is in the service of his antirealism about theories; the second point is central to his views on entity realism.

The realist believes that there are electrons. What is the experimental realist—who affirms that electrons exist but denies any commitment to theory realism—to make of the notion that *electron* gets its meaning from its role in theoretical laws? Hacking himself is happy enough to adopt in rough outline an account offered by Hilary Putnam according to which we must distinguish two aspects (among others) of the meaning of a natural kind term: its stereotype, which includes standardly accepted descriptions of the natural kind, and its actual reference—the thing it denotes or picks out.²⁷ Putnam allows that while we may succeed, at an initial baptismal naming or thereafter, in referring to an object under some stereotype, this does not entail that stereotypes “determine” referents in any strong way. As we frame clearer views and discover more about the thing or kind of thing, we can succeed in referring to the same thing or kind of thing even if the stereotype changes. Now, in the case of electrons, not only has the term *electron* undergone such stereotype changes, but moreover we presently have two incompatible stereotypes—the electron as wave and the electron as particle. Hacking implies that experimentalists can be confident that they are investigating the same thing, regardless of their commitments to this or that stereotype.

The notion of interfering or intervening forms the core of Hacking's experimental realism. We have seen that experimental realism recommends a shift away from theoretical representation, toward experimental manipulation. On Hacking's account, the very nature of scientific experimentation is to interfere in the course of nature. Indeed, if we are concerned with reality—with the entities making up the furniture of the natural world—then the fundamental grounds on which we form our beliefs are our causal interactions with such entities. They act on us, we act on them, and when we have understood their causal powers sufficiently well, we use them to interfere elsewhere in nature. Since this is a general fact about observables and unobservables alike, what justifies our belief in the reality of unobservables is no different in kind from what justifies our belief in observable objects. As Hacking puts it elsewhere:

Reality has to do with causation and our notions of reality are formed from our ability to change the world. . . . We shall count as real what we can use

to intervene in the world to affect something else, or what the world can use to affect us.²⁸

The implication of Hacking's stress on intervention is clear: we rely on the causal properties of objects—electrons, say—in developing instruments to generate various scientific results, and this very business of experimentation is incoherent unless we believe that such entities exist and have the causal powers they do. Thus, we do not infer the reality of electrons from experimental results; rather, we presuppose that electrons are real when we design and execute our experiments.

Hacking discusses at considerable length an example from experimental physics and then draws some morals from the example. He focuses on PEGGY II, an instrument used to fire polarized electrons at a deuterium target. The aim of the experiment was to observe parity violations in weak neutral interactions. Hacking concludes that constructing PEGGY II and executing the experiment described in his paper were largely non-theoretical affairs. The project began with a core set of beliefs about the properties of elementary particles and atoms, together with the discovery (in an unrelated experiment) that gallium arsenide (GaAs) crystals could serve as a source of polarized electrons. The inventors and designers of PEGGY II did not use a theory about electrons to explain or save any phenomena; rather, they used electrons to produce events that violate parity in weak neutral interactions. The experimenters were realists about electrons because they could intervene with electrons to create new phenomena. Thus, Hacking takes this to be a typical illustration of his thesis that it is experimental practice, not theorizing, that certifies scientific realism about unobservables.

9.9 | Resnik's Evaluation of Experimental Realism

Hacking is correct in claiming that the practice of scientific experimentation has been largely ignored in traditional disputes about scientific realism. As we have seen, it is rather questions about the relation between the world and our representations of the world (theories), about the nature of scientific inference and explanation, that have figured most centrally in the exchange between realists and antirealists. Hacking's project is to shift the debate from representation to intervention and to argue that, at the level of the entities themselves, the strongest argument for their existence is our ability to use them to control and create phenomena. Thus, Hacking is what one might call a *theory-free entity realist*. In "Hacking's Experimental Realism," David B. Resnik argues that this project fails in three respects: experimental realism is, at the end of the day, just another version of traditional success-of-science arguments; the experimental realist can

claim to have knowledge of unobservables only if she assumes that the theories describing them are at least approximately true; and experimentation is not nearly so theory-free as Hacking claims. We will examine these objections in turn.

According to Hacking, experimental science gives us the best grounds for believing in unobservable (theoretical) entities. If we use some part of the natural world to intervene—to control and create phenomena—then we are entitled to believe that our objects of intervention are real. Resnik reconstructs Hacking's argument as follows:

- 1 We are entitled to believe that a theoretical entity is real if and only if we can use that entity to do things to the world.
- 2 We can use some theoretical entities, e.g., electrons, to do things to the world, e.g., change the charges of niobium balls.
- 3 Hence, we are entitled to believe that some theoretical entities, e.g., electrons, are real. (1175)

While some have criticized Hacking's argument by attacking the first premise, Resnik is concerned with the grounds an experimental realist might have for asserting the second premise.

On Hacking's account, the assertion of (2)—that we can use theoretical entities, or unobservables, to do things to the world—is fundamentally a causal claim. How are we to support or justify causal claims? Hume argued that we do not deduce them from experience, but rather base them on inductive inferences: we observe regularities, or effects, in nature and infer causes from them according to the principle that similar effects must have similar causes. While Hume himself regarded such induction as ultimately unjustified (but the best we can do), many philosophers believe that we can justify causal claims inductively. Others believe that we can use abductive reasoning, inferring that *A* causes *B*, on the grounds that this best explains the available evidence. Now, in denying theory realism but accepting entity realism on the basis of causal arguments like the one above, Hacking is clearly committed to denying that we base our causal claims, inductively, on observed lawlike regularities.²⁹ Thus, while Hacking does not explicitly commit himself to any view about scientific reasoning, it would seem that his only alternative in the present case is abduction.

Resnik's objection is that Hacking's reliance on abductive reasoning places his project firmly in the camp of traditional success-of-science arguments for realism. As we have seen, these arguments conclude that we are justified in being scientific realists on the grounds that there is no other explanation for the success of science: unless the unobservables posited by our theories actually exist, the predictive success of science would be a miraculous coincidence. While these arguments focus largely on the

success of theories—on our ability to predict and explain—Hacking's argument focuses on the success of experiments—on our ability to construct and control instruments that use unobservables to create new phenomena. According to Resnik, Hacking is in the position of claiming that it would be an incredible coincidence that our instruments should work as they do, and let us affect the world as we do, if the unobservable entities we take ourselves to be manipulating in our experiments did not in fact exist.

Hacking rejects the traditional success-of-science arguments and denies that his own experimental realist argument is of this sort. Recall that the scientists involved in the PEGGY II experiment did not aim to explain any phenomena with electrons but rather simply to use them, and this is a general feature of experimental science. Clearly, Hacking is right in claiming that the experimentalists believe in the existence of electrons. Indeed, he may well be right in claiming that the experimenter "is *impressed* by entities that one can use to test conjectures," and that "the experimenter is *convinced* of the reality of entities . . . that . . . can be used to interfere elsewhere in nature" (1157; our emphases). But these are psychological facts about the experimenter. What justification Hacking's account provides for belief in those entities remains unclear. Our realist commitments to unobservable entities cannot be secured by dogmatic assertion. The only account available to Hacking, it seems, is the abductive one: such entities explain those very facts about our experiments that most impress us. And this has a further consequence. Entities by themselves explain nothing; rather, they explain only within the context of a theory that describes the phenomena to be explained. Thus, Hacking's theory-free entity realism appears to be jeopardized. This brings us to Resnik's second objection.

Hacking's theory-free entity realism is not incoherent: one can consistently believe in unobservables without believing that the theories about them are true. One can claim, for example, that the explanatory work done by laws of physics can be accomplished only if they are not literally true, while insisting that causal or experimental facts suffice for our beliefs in unobservables. The question is whether this "sufficing" in any way justifies our beliefs in unobservables. Resnik argues that, while Hacking's position is consistent, it is not reasonable, because it grants the experimental scientist belief in unobservables—and perhaps even true belief—without justifying that belief. And since justification is a necessary condition for knowledge, Hacking's experimental realism cannot provide knowledge of unobservables. Hacking may have explained why we hold the beliefs that we do, and why we hold them as strongly as we do, but he has not explained why those beliefs should be regarded as knowledge.

Recall again Putnam's view of meaning to which Hacking appeals in defending his experimental realism. As applied by Hacking, this view enables one to believe in unobservable entities belonging to a natural kind without believing in the theory about those entities, since we can suc-

cessfully refer to a natural kind (electrons, for example) quite independently of theoretically motivated changes in its stereotype. Resnik claims that, while this is cogent, it goes no distance toward providing an epistemological grounding for our belief in electrons. In particular, Putnam's view assists Hacking only if we know that the entities to which we refer are natural kinds (not artifacts or purely theoretical fictions, say), for unless we know that such entities are natural kinds, we cannot claim to know that we continue to refer successfully when our theories change. Thus, phlogiston was once posited as an unobservable, and phlogiston theory enjoyed some measure of success. But after many failures of theory, phlogiston was dismissed as fictitious. Similarly, early work in genetics posited the existence of hereditary units—genes—to explain various hereditary phenomena and formulated laws governing these entities. But these laws were shown to be inadequate, and contemporary molecular genetics has shown that the functions formerly attributed to genes are in fact performed by several different entities. In both cases, putative unobservable entities were shown to be artifacts. How, then, do we know that unobservables exist? How do we know that electrons are not like Mendelian genes or phlogiston? Claiming that we know electrons exist because we continue to refer to them despite changes in our theories will not do; this begs the question, for we cannot claim to refer successfully at all unless we are already justified in believing that the relevant entity is a natural kind.

The standard realist account of how we know that some unobservable entity is a natural kind is familiar: we know that some unobservable entity is a natural kind because we have well-confirmed, explanatorily successful, and approximately true theories positing their existence. We know that electrons are natural kinds because the theories positing them are well-confirmed and highly successful. Hacking, of course, rejects this but offers nothing in its place. Were he to claim that we know electrons are a natural kind because we can use them in scientific experiments, he would still have to explain why the experiments succeed in giving the results they do. The only hope of explaining these facts is by appealing to theories describing the causal goings-on in our experiments: scientists are justified in believing in electrons because theories about electrons explain the success of their experiments.

Hacking might respond to this argument by claiming that scientists can explain their experimental success using only fairly low-level generalizations of various sorts; they do not need full-blown theories—systematic accounts that include genuine scientific laws that are universal and nomically necessary. In reply, Resnik points out that adopting such a narrow view of theories entails that theories are confined to physics and perhaps chemistry, which seems an unacceptable restriction on the use of the term *theory*. Resnik then proceeds to his third and final objection to Hacking's experimental realism.

Even if the restrictive account of theories as full-blown systems of

universal scientific laws is granted, it nevertheless remains true that a working knowledge of such theories is essential to the design of experiments and the construction of measuring instruments and apparatus. The PEGGY II experiment, for example, was permeated with theory from beginning to end. Far from being mere recipe-following technicians, experimenters are trained in theory, communicate with theoreticians, and invent new ways of testing theories. Thus, claims Resnik, not only does Hacking's account fail to provide an epistemological grounding for the beliefs of experimenters, it also fails to describe their behavior as working scientists.

9.10 | Fine's Natural Ontological Attitude (NOA)

A familiar and much-debated argument for realism claims that the best explanation for the success of science is that scientific theories are true (or at least approximately true) and most (if not all) of their terms refer to real entities, properties, and processes. Rejecting traditional defenses of scientific realism at the level of theories, Hacking defends in its place an attenuated realism—a rather more deflationary, theory-free experimental realism solely about entities. Laudan and van Fraassen, too, reject traditional arguments for realism, van Fraassen going on to defend his own antirealist account of science, constructive empiricism. In “The Natural Ontological Attitude,” Arthur Fine agrees with Hacking and van Fraassen that the traditional arguments for realism fail. But Fine's own positive account—the natural ontological attitude of his paper's title—falls somewhere between Hacking's relaxed realism and van Fraassen's empiricist antirealism. According to Fine, the most compelling view of science and its theories is neither realist nor antirealist. Having examined in some detail responses to arguments for realism, we shall attend somewhat briefly to Fine's criticism of realism, focusing our attention rather more closely on this middle position that Fine calls the natural ontological attitude (NOA).

AGAINST REALISM

Fine's criticism of realism is directed along two separate lines: traditional abductive arguments for realism, and the place of realist and nonrealist attitudes in the development of successful twentieth-century scientific theories. What evaluation does Fine offer of success-of-science arguments for realism?

Fine objects to the overall strategy of realist arguments that aim to connect the methods of science with its overall success. He begins with a maxim—gleaned from the history of set theory and the foundations of

arithmetic—about the propriety of defending some theory *T* with argumentative resources taken from *T* itself. In the case of set theory, the formalist David Hilbert argued that, to the extent that the meaningfulness and consistency of Cantor's set theory was yet in question, only some non-set-theoretic consistency proof could warrant acceptance of Cantor's project. Hilbert thus advocated a more stringent method for establishing the consistency of set theory, a method that did not use arguments from set theory itself. Fine believes that Hilbert's attitude is the correct one: metatheoretic arguments in support of *T* should meet requirements more stringent than those placed on arguments at work in *T* itself.

Realists urge us to accept that realism offers the best explanation for why the methods of science lead to empirical success. In an objection reminiscent of Laudan's charge that inference-to-the-best-explanation arguments for realism beg the question, Fine notes that just as anyone who is suspicious of the consistency of set theory would scarcely be moved by set-theoretic arguments for its consistency, so also antirealists already skeptical about the connection between explanation and truth will fail to be moved by best-explanation arguments for realism. More generally, since realists and antirealists disagree about whether the realist's abductive arguments for the existence of unobservables are legitimate, even if realism is the best explanation for the success of science, this cannot warrant the conclusion that realism is true. Since it is the status of such best-explanation inferences that divides realist from antirealist, abductive defenses of realism at best beg the question.

Nor is there any pattern of argument more stringent than abductive ones, according to Fine, that would permit the realist to respect Hilbert's maxim. The only obvious candidate acceptable to all parties would be patterns of inductive argument leading to empirical generalizations. But empirical generalizations require observable connections between observables, and no such connection is available to the realist. The connection of relevance for realism is the relation of truth or approximate truth said to hold between theories and the world to which they correspond. To the extent that realism postulates a correspondence relation of truth between theories and the world, what it is committed to is unverifiable.

Fine's argument against the realist strategy rests on the rejection of abductive inferences. In taking this line, it would then seem incumbent on Fine to say what, if not abduction, ever does justify our belief in theories. Of course not all forms of induction will have been rejected. But now Fine and other nonrealists must offer some independent, properly stringent criterion by which to determine which inductive inferences are legitimate, without begging the question against the realist. In short, even if Fine is correct that realist arguments fail (which, of course, is not the same as showing that realism is false), his proposal leaves open whether any coherent nonrealist position can be sustained in place of realism.

In addition to criticizing traditional success-of-science arguments for

realism, Fine suggests that the history of science also tells against it. In particular, Fine argues that in two successful twentieth-century physical theories—relativity and quantum mechanics—we can find a refutation of the claim that realism is the only way to explain scientific progress. Both the special and general theories of relativity were the work of Einstein's early period, during which he espoused a broadly positivist methodology. And despite the fact that Einstein thereafter came to have rather more sympathy with realism, practitioners of relativity theory—working scientists—have not followed him in this, adopting instead an instrumental view of general relativity as a tool for solving large-scale gravitational problems. Likewise, Fine argues, the development of modern quantum theory was carried out in a largely nonrealist environment, both Heisenberg and the later Schrödinger expressly avoiding referential commitments to an unobservable reality. Indeed, Bohr is reported to have feared that Einstein's (later) realism, if taken seriously, would thwart the progress of the new quantum physics. By Fine's reckoning, the Bohr-Einstein battle, far from a tempest in a teapot, was in large measure the result of conscious efforts to preserve an increasingly successful theory from stagnation in the hands of realists.

It is, of course, no part of realism that scientific progress depends in any crucial way on *believing* that realism is true. A theory is successful, according to the realist, because it gets the world right, not because of anyone's belief that such a relation does or does not (can or cannot) hold between theories and the world. Nor is it clear how confident Fine or anyone else can be about how differently twentieth-century science would have looked had its founders viewed realism more congenially.³⁰ Nevertheless, it is undeniable that the two most important and successful theories of the twentieth century—relativity and quantum mechanics—have progressed in a largely positivist and instrumentalist environment.

THE NONREALIST “CORE”: NOA

Despite the nonrealist origins of quantum mechanics, Fine notes that contemporary particle physicists adopt models predicting the existence of new, as-yet-unobserved particles, judging these as discoveries when later confirmed in various experiments. Fine thinks that this talk of “discoveries” reflects not a return to realism but instead a consistent form of nonrealism that is equally at odds with antirealism. Thus, his task in the latter part of his paper is to develop a viable form of nonrealism.

Fine doubts that realists come to adopt their view of science on the basis of abductive reasoning. Rather, he suggests, they engage in a somewhat more homely line of argument of the following sort: “Well, on the whole I trust the evidence of my senses with respect to garden-variety objects. And on the whole I trust the checks and safeguards of working scientists, too. If they tell me that there are atoms or quarks or the like,

then I trust them and accept that there are such things. But I can accept the confirmed results of science, in just the same way that I accept the evidence of my senses, only if I am a realist about science. So I should be one."

Now, by Fine's lights, to accept scientific results in the same way that we accept the evidence of our senses is simply to accept them alike as true. But the antirealist could follow the homely line of reasoning sketched above to this extent: nothing about antirealism would incline him either to reject the results of science or to accord the confirmed propositions of science some secondary status distinct from that afforded the garden-variety truths of everyday life. So it is not at all clear that accepting the results of science as true has to bring realism with it.

Thus, Fine urges us to see that both the realist and antirealist alike must in the main toe the homely line above, that is, both must accept the confirmed results of science as quite on a par with other more familiar truths. Fine calls this acceptance of scientific truths the "core position." What distinguishes the realist from the antirealist is what they add to the core position. Antirealists may add on a pragmatic or instrumental or conventionalist analysis of the notion of truth, or they may add on some broadly epistemological construal of certain concepts, as would idealists, phenomenologists, and other sorts of empiricist, and so on. Realists, according to Fine, simply add on a desk-thumping, foot-stamping emphasis—"Really!" When the realist and antirealist accept the verdict of science that there are electrons, the realist adds the emphatic "There really are electrons, really!" (1199) in part to indicate a rejection of competing, antirealist additions to the core and in part to express the robust sense of truth generally understood to accompany realism—namely, truth as correspondence with the world. (Fine reckons this latter aspect of the realist addition to be merely a superficial decoration of no more rational force than foot-stamping itself.)

From this contrast between what the realist and antirealist add to the core position, Fine urges us to see that an attractive third alternative emerges: the core position itself. Neither a form of realism nor antirealism but at the core of both, this intermediary view seems to be the most compelling and sensible attitude to adopt toward science. Fine calls this position the "natural ontological attitude" (NOA, pronounced *Noah*).

Fine thinks that the adequacy of NOA as a philosophical view of science can be appreciated by seeing what it has to say about ontology and the methods of science—two issues figuring crucially in most assessments of realism and its competitors. Focusing here on the first of these, consider that NOA tells us to accept the results of science as true. To do so, Fine claims, is to treat truth in the normal referential way: statements of science are true only if appropriate entities (individuals, events, properties, etc.) exist and stand in the operative referential relations. In adopt-

ing this stance, NOA is not thereby committed to the thesis that long-run scientific success inevitably brings us closer to the truth. A NOAer (pronounced *knower*), who in some particular scientific context or tradition believes in the entities referred to by her theories, can consistently allow that there is nothing progressive about changes in the scientific context or tradition and that Kuhnian paradigm shifts might well permit no stability of reference at all. So, too, can the NOAer resist believing that talk of reference entails belief in some objective external world. According to Fine, those who engage in such talk only pretend that we can step outside the game of reference and science to judge from some higher view what science is about, to mark off a world of external objects that are the referents of our scientific terms. But, Fine insists, we ourselves are in the world, among the objects of science.

Fine suggests that NOA can help to clarify particular connections between scientific method and the realist success-of-science accounts judged earlier to be question begging. But the main focus of Fine's proposal is not NOA's ability to clarify and explain. After all, the core position that NOA represents is at the heart of both realism and antirealism, and if the realist or antirealist additions do not nullify NOA's appeal to the truth of scientific claims, its ability to explain and clarify will to that extent be equally available to its competitors. What Fine wants us to appreciate above all is how minimal an adequate philosophy of science can be. According to NOA, the legitimate aspects of any addition made by realism or antirealism are already implicit in the core belief that the confirmed truths of science are on a par with garden-variety truths, all alike and equally true. NOA brings with it no inflation of the concept of truth, no theory or analysis or picture of truth as correspondence or anything else. In addition to accepting the standard referential semantics and broadly Tarskian conception of truth, the NOAer accepts a standard epistemology by which judgments of truth are based on familiar perceptual and confirmational relations.³¹ Unlike either realism or antirealism, NOA makes no pretense to having the resources for settling many disagreements over what is true or even whether best-explanation inferences are truth preserving. And that is as it should be, according to Fine: there are no such resources to be had. NOA's minimalism is thus among its chief virtues. Science does not need what either realism or various versions of antirealism suppose it to need. A global interpretation, a philosophical version of truth to explain the success of science, a grand story purporting to render the significance and goals of science more intelligible—all these things are unnecessary. So the minimalism of NOA reflects the attitude that science can be taken on its own terms, an adequate philosophy of science requiring little more than a homely line of reasoning about the history and practice of science just as we find it. The philosophical theories, the interpretations, the global pictures, the truthmongering "isms," all are idle

trappings, unnecessary, unwarranted, perhaps even unintelligible bits of decoration appended to philosophy of science. Nothing else is required, nothing else is legitimate, beyond the natural ontological attitude.

9.11 | Evaluating NOA: Musgrave's Realist Reply to Fine

In "NOA's Ark—Fine for Realism," Alan Musgrave asks, What exactly is there in NOA that runs counter to scientific realism? How nonrealist is Fine's proposal?

Musgrave notes at the outset that, insofar as the debate between realists and antirealists centers largely on the question of truth, NOA seems less a core position that realists and antirealists would then embellish than a position that realists would accept and antirealists reject. The core position would have us accept the claims of science as true. This may be fine for proponents of realism, but not for its naysayers: positivists who deny the existence of theoretical entities judge the theoretical claims of science to be false; instrumentalists reckon scientific theories as neither true nor false; and empiricists like van Fraassen insist that theories should be accepted as empirically adequate, not as true. At times, Fine seems to think that these and other versions of antirealism import a peculiar notion of truth into their philosophies of science. But were that so, antirealists could scarcely accept the core position's directive to treat the results of science with the same homely conception of truth that we apply to garden-variety beliefs.

Perhaps there is no such homely conception of truth, and the core position itself leaves open what sense realists and antirealists should bring to the word *true*. But Musgrave doubts that this will do justice to Fine's intentions. For without any sense attached to *true*, there is no determinate core position to which realists and antirealists alike could then make additions. Moreover, it looks as if the core position (NOA) attaches a quite definite sense to *true*. Fine says that "we are to treat truth in the usual referential way, so that a sentence (or statement) is true just in case the entities referred to stand in the referred-to relations," adding that "NOA recognizes in 'truth' a concept already in use and agrees to abide by the standard rules of usage, [these rules involving] a Davidsonian-Tarskian referential semantics" (1203). In passages like these, Musgrave urges that Fine has already embraced a quite deep philosophical analysis of truth, one that realists would (but antirealists would not) be prepared to apply across the board, to homely and scientific truths alike.

Fine intends NOA to be a nonrealist philosophy of science. What is it that realism adds to the core position that distinguishes the realist from the NOAer (that is, the person who accepts Fine's NOA)? Drawing partly on other work of Fine's, Musgrave isolates three items: first, the realist

shouts, emphasizing “really!” whereas the NOAer does not; second, the realist makes a good deal of fuss over certain philosophical slogans that the NOAer avoids; and third, the realist is committed to a metaphysical picture that the NOAer is not. Let us consider these in turn.

The first putative difference is judged by Musgrave to be the shallowest: if the NOAer is a realist who avoids shouting and foot-stamping, the difference is an unphilosophical one, easily dissolved by behavior modification. But, of course, that is too facile: the difference is meant to run more deeply than that. For there is, Fine says, a positive function to the realist’s emphasis—namely, an expression of a robust sense of truth, the full-blown version of which involves truth as correspondence to the world. And this brings us to the slogans.

Musgrave’s second point is that, according to Fine, realists endorse slogans such as “Truth is correspondence with reality,” whereas NOAers will have no truck with correspondence or anything like it. For his own part as a realist, Musgrave wonders what more there can be to the slogan “Truth is correspondence with reality” that is not already expressed by Fine’s view that “a sentence (or statement) is true just in case the entities referred to stand in the referred-to relations” (1200). Not all philosophers would agree with Musgrave that Tarski’s account of truth is just a codification of the commonsense correspondence theory, though exactly what more there is to correspondence than Tarski’s theory plus a standard referential semantics is a point of debate. In any case, it is far from clear that scientific realism requires anything more than Tarski offers. As Musgrave sees it, the realist need not suppose, for example, that in addition to a Tarskian semantics one needs a story about the essence of truth. Indeed, Musgrave agrees with Tarski (and Fine) that there is no property shared by all truths, in virtue of which they constitute a natural kind. If that is correct, then, despite what Fine says, NOA is a realist account of the claims of science.

Musgrave also responds to the concern that, if Tarski’s theory supplies everything the realist needs, then it threatens to make us realists about too much. If a Tarskian referential semantics applies across the board (to homely claims, theoretical-scientific claims, and everything in between), then such a liberal application of Tarski’s Convention T would seem to recommend realism about numbers, moral claims, and the “creeps,” no less than it recommends realism about the homely moon and the scientific electron. All five instances (below) of Convention T should be treated in exactly the same way:

- 1 The statement ‘There is a full moon tonight’ is true if and only if there is a full moon tonight.
- 2 The statement ‘Electrons are negatively charged’ is true if and only if electrons are negatively charged.

- 3 The statement 'Two plus two equals four' is true if and only if two plus two equals four.
- 4 The statement 'Eating people is wrong' is true if and only if eating people is wrong.
- 5 The statement 'Ronald Reagan gives me the creeps' is true if and only if Ronald Reagan gives me the creeps. (1214)

On the Tarskian referential semantics that, according to Musgrave, is part of scientific realism, it seems as though we should be realists about numbers and the "creeps" as well as about some remarkably average family and partial children: the statement "The average American family has 2.3 children" is true if and only if the average American family has 2.3 children.

To avoid such commitments, Musgrave claims that Tarski's theory by itself yields none of these realisms. This is surely right. There is, for example, nothing in Tarski's theory that requires us to take every sentence at face value. If one is charitable about the utterance figuring in (5) above, but is nevertheless keen to avoid a realist commitment to the creeps, one can allow that the relevant sentence is a true but idiomatic expression of a fact about the way Musgrave felt about Reagan—this bringing with it at most a commitment to Reagan's existence and to something Reagan did. Alternatively and less charitably, one can take the sentence in (5) at face value and judge that it is, strictly speaking, false. Similarly, the antirealist about numbers can take "Two plus two equals four" at face value and say that it is strictly false, since there are no numbers answering to the words *two* and *four*. The same options apply to the homely and scientific cases of (1) and (2) respectively.

In short, Tarski's account brings no particular realistic commitments with it. What it does do is make realism possible. Musgrave's concern is to show that Fine's account of NOA renders realism about the claims of science actual. For NOA enjoins us to take the claims of science at face value, to apply the Tarskian referential semantics, and to accept the relevant claims as true. And that is scientific realism.

One might construe NOA as claiming that a face-value reading of sentences is already a departure from the core position, since something has been added to the minimal stance. And here, perhaps, there is room for distinguishing the NOAer from the realist. Where the realist accepts the claims of science (and homely run-of-the-mill claims) as true, reads *true* in the Tarskian way, and takes the claims at face value, the NOAer refrains (on the present reading) from doing the last of these, thus leaving open the issue of commitment to particular entities by virtue of leaving open the issue of what reading (literal and face-value or idiomatic and misleading) to give the claims. As Musgrave notes, this reading renders NOA a far cry from any sort of realism: for it characterizes the NOAer as

one who is not committed to anything at all—not to electrons, not to the moon, not to physical objects or to other people.

This interpretation of NOA, as an extreme form of philosophical minimalism, eventually reemerges in the context of the third respect Musgrave notes in which Fine reckons NOA to depart from realism: realism offers a metaphysical picture that NOA rejects. In a later paper entitled “Unnatural Attitudes: Realist and Instrumentalist Attachments to Science”³² (published two years after the original NOA paper included in our volume), Fine characterizes the realist metaphysic as one according to which science is about something, something out there, external to us—in short, the World. Realists (Musgrave among them) concur with this picture and wonder what it is that Fine and the NOAer reject about it, and what NOA puts in its place. As for what NOA rejects in the realist picture, Fine’s misgivings concern what he calls the problem of access: How can we render intelligible the realist’s claim of having access to an objective, independent external world? For whatever we observe—or, more properly, whatever we causally interact with—is not independent of us, and whatever information we retrieve from such interaction is information about interacted-with things. Faced with these problems of reciprocity and contamination, Fine contends that the realist can offer no intelligible account of our supposed access to an independent and objective world.

As Musgrave notes, if there is a genuine difficulty for the realist here, it will be perfectly general, infecting both commonsense realism (about the moon, say) and scientific realism (about electrons). So he sticks to commonsense realism: “Somebody says ‘There is a full moon tonight,’ and I look up into the night sky and ascertain that the statement is true” (1217). This captures the realist notion that sentences express something true if they represent the world to be a certain way when the world is that way. Now when Musgrave claims that there is an independent and objective world about which our sentences can express truths or falsehoods, what he means is not that things in the world are causally insulated, but rather that the existence and nature of these things is not dependent upon any perceptual or other mental goings-on in us. The moon, like an electron, is not something we create by looking at it or thinking of it or speaking about it; indeed, there is some *it* that we see, some object about which we speak and think. By Musgrave’s lights, the putative problem of reciprocity is no problem at all. To say that the moon or electrons fail to be independent because we interact with them is to invoke a strange notion of independence that no realist would accept.

What of the problem of contamination? Fine’s suggestion is that when we interact with the moon, we gain information not about an objective moon out there in the external world, as it is in itself, but about an interacted-with-moon. Now if this is to show that the interacted-with-moon is not objective, not out there in the external world, then presumably the interacted-with-moon is in some sense subjective and inside our heads.

But what sense can be offered and defended here? Musgrave doubts that Fine wants to return to the dusty old view that we do not see external objects (such as moons) but instead merely have commerce with (moon-ish) sense data.³³ The nonrealist NOAer might urge instead that the interacted-with-moon, the moon-as-seen-by-us, fails to be objective in this sense: insofar as my judging that the moon is full requires, and so depends upon, an application of some moon-concept I happen to possess, this judgment and its truth does not concern any purely external-world moon as it is in itself but rather some world partly of my making, a world-as-conceptualized-by-me.

Some philosophers unsympathetic with realism will speak in this way. Musgrave makes two points about the philosophical view such talk represents.³⁴ First, this sort of conceptual idealism leads quickly to conceptual relativism. Since concepts vary (from person to person and from culture to culture) and change (from olden days to current enlightened days), there is no single world-as-conceived-by- Conceived by whom? When? We have, apparently, a plethora of such worlds, all different. Now this point of Musgrave's is not yet an objection: nothing he says indicates what is wrong with relativism, and some nonrealists will find it congenial. But Musgrave offers a second point that is intended to be more damaging—a point concerning not the plethora of such conceived worlds, or the many instances of the moon-as-conceived-by- . . . , but instead the relation of the moon-as-conceived-by-us, say, to the moon-in-itself. What exactly is the relation between them? Are they numerically identical, one and the same moon? If so, there is no point in speaking of a distinction between the moon-as-conceived-by-us and the moon-in-itself: let us simply speak of the moon and be done with it. But if the moon-as-conceived-by-us is not the same thing as the moon-in-itself, then there is some property the one has that the other lacks, by virtue of which they are distinct. And yet to know this is to know something about the moon-in-itself. But this, according to the contamination problem, is impossible, all knowledge being confined to the moon-as-conceived-by-us.³⁵

What is the upshot of all this, according to Musgrave? Clearly, the various (and sometimes tedious) routes we have been pursuing most recently in following out Fine's rejection of the realist metaphysic represent something well beyond a natural ontological attitude or a philosophically minimalist one. The accounts of reciprocity and contaminated information, the conceptual idealism, the relativism—all of these are prime examples of philosophical theorizing, precisely the sort of thing Fine had enjoined us *not* to add to the homely core position that is NOA. It rather looks as if Fine rejects the realist's metaphysic only by abandoning the minimal core, by inflating NOA beyond its proper boundaries. By returning to the natural core—judging the claims of science (like those of common sense) to be true, applying the Tarskian semantics, and taking these claims at face value—one has accepted the realist metaphysic.

Musgrave adds one final note, returning us to the extreme philosophical minimalist construal of NOA (broached in our earlier discussion of realist slogans). This philosophical minimalism is an option the NOAer might take. The thoroughly unphilosophical natural ontological attitude would have one accept parts of science as true, without this implying anything at all about the realist's claims of the objectivity and independence of electrons or the moon. By leaving it an open question whether the accepted statements are to be taken at face value or not, it is left open how these statements are to be interpreted and what one is ontologically committed to when one accepts them. This version of NOA—Fine's, perhaps—is not a philosophical view at all. Moreover, since NOA enjoins us to treat the claims of science and the claims of everyday life on a par, pretty much all sentences whatsoever (ordinary, scientific, philosophical) will be “open” in the sense just described. Musgrave concludes that if NOA is to have any content at all, it must at least give the homely, run-of-the-mill claims a realist, face-value hearing. And then, of course, given NOA's insistence that the claims of science are on a par with the run-of-the-mill claims, NOA emerges once more as “fine for realism.”

9.12 | Summary

Are scientific theories literally true, and do the entities apparently postulated by them in fact exist? Logical empiricists approached such questions by distinguishing observational terms and observables, on the one hand, from theoretical terms and unobservables on the other. About the observational terms and observables of accepted scientific theories, the logical empiricists espoused a form of realism: claims involving only observational terms were judged to be true and genuinely referential, and observables were said to exist. But they were antirealists about theoretical terms and unobservables. Most logical empiricists were committed to the meaningfulness of theoretical terms and the truth of the theoretical claims in which they figure without also accepting that theoretical terms refer or that unobservables exist.

Grover Maxwell argues that the antirealism of traditional empiricism cannot be sustained. Confronted with scientific advances in which objects previously judged unobservable come to be seen with the help of scientific instruments (such as microscopes), some empiricists had claimed that strictly speaking we see, not the objects, but only images in our instruments. Maxwell argues that such a view entails that strictly speaking we do not see ordinary objects through binoculars or eyeglasses or even through windowpanes. Since we are realists about observables and since there is no nonarbitrary line to be drawn between the observable and the unobservable, we have no grounds for being antirealists about unobserv-

ables. Other empiricists restricted their antirealism to what is entailed by accepted theories to be *in principle* unobservable. Maxwell argues that there are no conceptual or a priori limits on what is observable: any object is a candidate for being observable in principle: imagine having electron-microscopic eyes or X-ray vision. Finally, some empiricists insisted that all claims about physical objects be translatable into claims about immediately perceptible sense data. While this view finds few contemporary adherents, it draws our attention to what is widely accepted—that there must be some fundamental observational base on which confirmation in empirical science ultimately rests. Maxwell argues that the question of what things count as observable (and hence qualify as members of the observational base) is ultimately a theoretical question and that the observational-theoretical distinction will be drawn on the basis of our current state of theoretical knowledge, our physiology, and the instruments we happen to have available to us. But this shows that the distinction has no ontological significance whatsoever.

To show that the logical empiricists failed to make their antirealist case is not yet to show that scientific realism is true: other forms of antirealism may be waiting in the wings. Bas van Fraassen defends a kind of antirealism that is at odds both with logical empiricism and with scientific realism. As van Fraassen understands scientific realism, it is the view that science aims to give us literally true theories about the world and that to accept a theory is to accept it as true. Antirealists deny this. Rather than offering a theory as true, the scientist can claim for the theory some other virtue instead, remaining agnostic about its truth. The virtue that van Fraassen claims a proper empiricist will require in accepting a scientific theory is empirical adequacy: a theory is empirically adequate if it “saves the phenomena”—if what it says about observable objects and events is true. According to van Fraassen’s constructive empiricism, then, the aim of science is to offer theories that are empirically adequate, and when we accept a theory, we accept it, not as true, but as empirically adequate.

How well can the constructive empiricist defend this distinctive breed of antirealism against the realist arguments of Maxwell and others? Van Fraassen’s notion of empirical adequacy entails that the observable-unobservable distinction can be made in an intelligible and defensible way. Maxwell argued that the distinction cannot be made out, since there is no nonarbitrary line to be drawn along the continuum from observable objects (seen through the air or through a window) to unobservable objects (detectable only with the aid of instruments). Van Fraassen argues that this last claim entails only that *observable* is a vague predicate and that we can still use it safely in clear cases where it applies or does not apply. Van Fraassen regards detecting particles in a cloud chamber as a clear case where *observable* does not apply: we observe the vapor trails such particles leave behind, but we do not observe the particles themselves. Van Fraassen also denies Maxwell’s claim that the phrase *in principle unobservable* fails

to mark off any class of objects whatsoever. Just as we do not judge that the Empire State Building is portable (claiming that there are no conceptual or a priori constraints against the possibility of a giant moving it), so we judge what is unobservable in science on the basis of the natural limitations of humans: given these limitations, some objects *are* unobservable.

Van Fraassen rejects a number of other realist arguments. For example, realists often claim that we are justified in believing in the existence of unobservable objects by an inference to the best explanation: given some class of observable phenomena, our best explanation for what we observe requires us to believe in (posit the existence of) unobservable entities. Van Fraassen agrees that explanation is a legitimate criterion to use in arguing for what should be believed, but he denies that this pattern of inference commits us to realism. Since the claim that we follow such an inferential rule is an empirical one, it can be confronted with a rival hypothesis—van Fraassen's alternative hypothesis is that we are willing to accept that the theory which best explains the evidence is (not true, but) empirically adequate. Van Fraassen also rejects a number of other arguments for realism (from Smart, Sellars, Putnam, and others) intended to show that only a realist construal of theories and unobservables can explain the empirical success of our scientific theories. Perhaps the most influential of such arguments is the so-called Ultimate Argument: if we do not believe that our theories are true and that many of their terms really do refer to unobservable entities, then we can only reckon the success of our theories a miracle. To this "scientific" defense of realism (a defense that takes seriously the notion that even our philosophical theories must meet the demands of adequate explanation), van Fraassen replies with an alternative Darwinian explanation for the success of science. Just as many species of organisms have struggled for survival, so have many past theories; and just as organisms that failed to adapt to their environment no longer exist, so theories that are not empirically adequate no longer exist. There is no miracle in either case. Present-day theories are largely successful because scientists select for empirical adequacy.

Alan Musgrave has offered several criticisms of van Fraassen's constructive empiricism. In going beyond the observable to affirm the existence of unobservable entities, realism seems to be more susceptible than antirealism to skeptical arguments. Musgrave argues that the realist takes no greater risk of being proved wrong on empirical grounds than does the antirealist, given that empirical adequacy requires a theory to save all observable (not simply observed) phenomena. And while van Fraassen believes that the epistemically responsible course to take is mere acceptance of a theory, not belief in its truth, Musgrave claims that, since the risk associated with either option is the same, one should opt for the course offering the chance of the greatest gains. However the details of this issue might be viewed, the observable-unobservable distinction lies at the heart of the debate. Van Fraassen believes that even if the term *observable* is

vague, there are clear cases where it does not apply (to particles in the cloud chamber, for example) and, moreover, that the natural limitations of humans make relatively clear what it means to say that something is in principle unobservable. Musgrave notes that such limitations vary from person to person and are a function of our evolutionary history. So he doubts that the distinction can have the philosophical significance the antirealist ascribes to it. Van Fraassen claims that the distinction need only serve the epistemological role of specifying the extent of one's beliefs: believe to be true only what a theory says about what can be observed. Musgrave thinks that this requirement is still too strong. Given any plausible account of evidential support, it is at least possible that there will be better evidence for some claim involving unobservables than for a claim restricted solely to observables. Moreover, he finds it unreasonable to suppose that a scientist can agree that we detect (not observe) a particle in the cloud chamber while yet not believing that the particle actually exists. Finally, Musgrave argues that van Fraassen's Darwinian response to the Ultimate Argument fails. Just as the scientist can ask not only why a particular mouse runs from the cat but also why mice in general have cat-fleeing behavior, so also in the case of the success of science, one can ask not only why some particular theory is successful, but also why existing theories in general are successful. Van Fraassen has answered the second question: existing theories are successful because unsuccessful theories do not survive. But this does not answer the first question about why any particular theory is successful, and Musgrave believes that the antirealist has no answer to this question. In the end, the constructive empiricist seems forced to say that the success of any particular theory about the world is a miracle.

The Ultimate Argument is a popular defense of realism—an inference-to-the-best-explanation argument applied to science itself. According to such abductive arguments, the success of science is inexplicable unless its theories have the virtues typically attributed to them by realists. Chief among these virtues are the approximate truth of scientific theories and the genuine reference of their theoretical terms. Larry Laudan argues that the connection between these virtues and the success of science is much weaker than realists suppose. With regard to reference, realists typically claim that reference explains success while success warrants a presumption of genuine reference. Laudan argues, first, that the history of science supplies numerous cases of unsuccessful theories about objects we believe to exist: reference is no guarantee of success. Likewise, Laudan argues that many historical theories have been successful without being genuinely referential: the success of a theory creates no presumption of genuine reference. With regard to truth, realists typically claim that theories, if strictly speaking false, are nevertheless close to the truth. Laudan argues that realists have not clarified the notion of approximate truth, or verisimilitude, sufficiently for us to evaluate the thesis that approximate

truth entails explanatory success. Does the success of our theories warrant belief that they are true or approximately true? Laudan thinks that realists would not want to claim that a theory was approximately true if its theoretical terms failed to refer. This, together with historical cases of successful but nonreferring theories, shows once again that the connection between success and truth is not as close as the realists have assumed. Finally, one might question the fundamental logic of traditional success-of-science arguments. Antirealists have long denied that, on the basis of its yielding true consequences, some theory could reasonably be judged true, for such an inference would commit the fallacy of affirming the consequent. Laudan notes that, while success-of-science arguments are not about any theory in particular but about the truth of realism itself, nevertheless they are of the same form: realism can reasonably be judged true because it yields true consequences (science is successful). And such arguments simply beg the question against the antirealist, who, having rejected the application of this form of argument to particular scientific theories, cannot be expected to accept its application to science as a whole.

Traditional attempts to defend realism lean heavily on the notions of verisimilitude (approximate truth) and success. James Robert Brown explores realist and antirealist arguments in the light of more precise characterizations of these central notions. He observes that successful scientific theories have three main features: they explain and unify a wide variety of known phenomena (empirical adequacy); later theories are more adequate than earlier ones (increasing adequacy); and they make novel predictions that turn out to be true more often than mere guessing or coincidence would allow (novel predictions). Brown argues that van Fraassen's Darwinian explanation of the success of science fails to explain the third feature of successful theories: novel predictions. Moreover, since the relevant criteria of rational theory choice are not, as van Fraassen supposes, exhausted by purely empirical considerations, it is unclear how his account can capture adequately the first two aspects of success.

Following Laudan, Brown also rejects the realist's claim that genuine reference by theoretical terms explains why our theories are predictively successful. What about truthlikeness? Here Brown offers a much more detailed account of truthlikeness than is usually given, drawing on an analysis of verisimilitude by William Newton-Smith. On this account, one theory has greater verisimilitude than another if and only if it says more about the world (has greater content) and more of what it says about the world is true (has more true content). This analysis of truthlikeness would provide the crucial connection, needed by realists, between verisimilitude and likelihood of observational success. But Brown argues that the analysis ultimately fails. For not only does it not explain why present theories get so much right; it also does not account for a number of cases where one theory has greater verisimilitude than another even though its content is smaller.

Can we explain the success of science? Brown argues that a rejection of abductive, success-of-science arguments for realism cannot be based on a rejection of inductive argument generally: even the antirealist will accept ampliative inferences of many sorts. But the antirealist's Darwinian answer appears to fail, and the realist has yet to make good on the notion of approximate truth and its relation (along with reference) to success. Brown suggests that the difficulty lies in our having the wrong picture of explanation. He thinks that explanations need not be arguments (contrary to what Hempel and others have insisted) and that they need not specify sufficient conditions for what they explain (contrary to Hempel's D-N model). Brown gives narrative explanations and explanations based on relations of statistical relevance as examples that meet neither of these two conditions yet still, allegedly, explain. Thus, he argues, we can allow that truth does explain success (in some weak sense of *explanation*), without having to claim either that we know our theories to be true or that their truth guarantees their success.

According to Brown's relatively weak connection between realism and the success of science, realism may explain success, but we can no longer expect success to justify our belief in realism. What sort of considerations might one then bring to an argument for realism? Ian Hacking proposes that we divorce our commitment to the existence of unobservables from our belief that scientific theories are true. According to his experimental realism, our warrant for believing that electrons exist consists in our ability to manipulate them and to use them to intervene elsewhere in nature. This ability is based on our knowledge of their causal properties and entails no commitment to believing that our theories are true. Hacking adopts Putnam's view of meaning, according to which our theoretical terms can retain the same unobservable objects as referents during periods of scientific progress, even if the theoretical stereotype associated with these terms changes. In addition, Hacking emphasizes that we form beliefs about the existence and nature of objects on the basis of our causal interactions with them. But causal interaction and its role in belief formation is a general fact about observables and unobservables alike; so what grounds our beliefs about unobservables is no different in kind from what grounds our beliefs about observables. Thus, causal interaction, based on our knowledge of the causal properties of unobservables, is a crucial element in Hacking's experimental realism. Scientific experimentation grounds realism about unobservables because it involves causal interaction with them.

Hacking illustrates his view by describing in some detail the PEGGY II experiment—the design, construction, and use of a polarizing electron gun to observe parity violations in the scattering of polarized electrons. The PEGGY II example is meant to show that neither theory nor explanation is central to the practice of intervening. So the point is not to argue for realism by appealing to the explanatory success of theories that

posit unobservables. The defense of realism about unobservables rests on our ability to use and manipulate them.

While the attention Hacking has drawn to experimentation is important, David Resnik argues that his attempt to divorce realism about entities from realism about theories does not succeed. Moreover, Resnik objects, if Hacking's experimental realism is to succeed, it must inevitably rely on the same form of reasoning as the traditional success-of-science arguments for scientific realism. Given that causal judgments figure centrally in the notion of experimental intervention with unobservables, how are we to justify causal claims? Since Hacking denies realism about laws and theories, he cannot use causal laws (supported by inductive evidence) to justify his realist beliefs about the causal powers of objects. Since Hacking cannot use inductive inferences to justify causal claims about unobservables, he is left with only abductive arguments to play the justificatory role. While traditional realists focus on the theoretical successes of prediction and explanation, Hacking focuses on experimental success—on our ability to construct apparatus and instruments to create, measure, and control new phenomena. Hacking is thus in the position of claiming, with the traditional realist, that unless we believe in unobservables, we should judge it a coincidence or a miracle that our experiments work as they do.

Moreover, Resnik argues that wedding realism about entities to anti-realism about theories, while coherent, is unreasonable. At best, it allows the experimental realist to account for the beliefs we have, without supplying any justification for them. This is because, given Hacking's deployment of Putnam's account of reference, our belief in the existence of electrons, say, can be justified only if we know that electrons are a natural kind. And according to Resnik, this requires an appeal to theory. In other words, electrons can be judged to play the roles they do in experimentation only if we accept as approximately true theories that posit their existence. Thus, despite its initial promise, Hacking's experimental realism does not appear to provide scientific realism with any new justifying arguments.

A fundamental sentiment motivating Hacking's proposal is a general rejection of realism about theories. Here Arthur Fine concurs: even if the truth of scientific theories is the best explanation for their success, this cannot warrant belief in scientific realism. For since the realist and anti-realist disagree over the epistemic value of abductive arguments, such arguments for realism beg the question against the antirealist.

But Fine's belief that no argument for realism is successful does not lead him to espouse any of the traditional forms of antirealism. Indeed, he thinks there is no need to endorse either realism or antirealism. According to Fine, to accept the confirmed results of science is to treat them just as we treat the evidence of our senses: we accept them as true, period. Surely the antirealist need not reject the results of science, nor give them some secondary ranking below the perceptual truths of everyday life. Both realist and antirealist alike must at least adopt this core position; what

separates them is what they add to this core. Antirealists add some particular analysis of truth, or some empiricist treatment of certain concepts; realists add an emphatic “Really!” to the claims of science, expressing some more robust notion of truth as correspondence with the world. But Fine insists that an attractive third position between the realist and antirealist philosophies of science is the core position itself—a minimal, commonsense, nonrealist attitude toward science, the natural ontological attitude (NOA).

According to Fine, to accept the results of science as true is simply to treat truth in the normal, referential way, without any further commitment to the long-run progress of science as converging on the truth. Proponents of NOA can consistently believe in the entities referred to by their theories while also allowing that there is no referential stability across scientific traditions or paradigms. Nor does NOA bring with it any inflated notions of truth or any resources for settling particular disagreements over the concept of truth. The commonsense minimalism of NOA brings to philosophy of science all that it needs. Science can be accepted just as the homely core position enjoins us to accept it, without any philosophical trappings, interpretations, or fancy pictures.

Musgrave doubts that on its own, NOA is any sort of nonrealism at all. In the first place, many antirealists deny that we should take the claims of science at face value and accept them as true. To accept the claims of science as being true in just the same homely sense that we judge our ordinary beliefs to be true is the position of the realist, not of the antirealist. In the second place, if there is no common, homely conception of truth about which realists and antirealists can agree, then it is unclear what NOA is recommending when it tells all of us—realists and antirealists alike—to accept the claims of science as true. Indeed, Musgrave suggests that insofar as Fine associates NOA with a standard referential semantics, he has already adopted a philosophical analysis of truth, and it is exactly the analysis that realists would accept.

What is it, then, that realism adds to the core position to distinguish it from NOA? The foot-stamping “Really” is nothing philosophically deep, and Musgrave urges that there is nothing more to a correspondence theory of truth than a broadly Tarskian referential semantics, which the core position endorses. Fine insists that realism brings with it metaphysical commitments that NOA does not. In particular, Fine characterizes realism as embracing the view that science is about an objective, independent external world. The difficulty with such a view, according to Fine, is that we cannot explain how access to such a world is possible: perceptual interaction with objects renders them no longer independent, and whatever information we get from interacting with them will be information about interacted-with things, not objective things after all. Musgrave argues that Fine’s notion of independence is nothing the realist would recognize and that Fine’s worry about objectivity is in fact a concession to conceptual

relativism. Whatever the value of Fine's arguments against the metaphysical commitments of realism, they are clearly philosophical in nature. To this extent, Fine seems able to defend NOA against realism only by adding to the core position just what NOA requires us to omit—philosophical pictures, interpretations, and so on. In the end, Musgrave suggests that if NOA is not to emerge as philosophically empty, we must construe its commitment to the truth of scientific claims as one that is fully congenial to realism.

■ | Notes

1. Logical positivism was the official doctrine of the Vienna Circle or *Wiener Kreis*, a group of philosophers, mathematicians, and scientists led by Moritz Schlick at the University of Vienna in the 1920s and early 1930s. By the mid-1930s it had evolved into logical empiricism and spread to Britain and America as philosophers of science fled Nazi persecution in Europe. (Schlick was murdered by a deranged student in 1936.) Impressed by the success and rigor of the sciences, and unhappy with the verbose obscurity of most of European academic philosophy, the logical positivists insisted that empirical claims are cognitively meaningful only if they can be conclusively verified by observation. Metaphysics, ethics, and theology were among the traditional branches of philosophy thus reckoned to be cognitively meaningless. By the mid-1930s, leading positivists such as Rudolf Carnap realized that the verifiability criterion of meaning was too strict, since it would rule out as meaningless, not only metaphysics and theology, but also most of science. (As Popper pointed out, no universal generalization, and hence no law of nature, could be conclusively verified.) Similarly, it was realized that many scientific concepts could not be given complete, operational definitions solely in terms of observables, as the positivists had hoped. Thus, the verifiability principle was abandoned and the transition to logical empiricism was begun. Instead of verifiability, the logical empiricists took confirmability as their criterion of cognitive significance, and instead of trying to give complete operational definitions for theoretical concepts, they turned their attention to the analysis of the meaning of theoretical statements in terms of their observational consequences. Logical empiricism was a tremendously influential movement in the philosophy of science, although its popularity began to decline in the 1960s as it confronted a number of tough problems and objections.

Hempel's covering law model of explanation and Nagel's analysis of the logic of reduction are good examples of the logical empiricist approach to understanding science. For many years, logical empiricism was "the Received View" in philosophy of science, and it continues to exert a considerable influence. For a comprehensive account of the rise and fall of logical empiricism, see Frederick Suppe, "The Search for Philosophical Understanding of Scientific Theories," in *The Structure of Scientific Theories*, 2d ed., ed. F. Suppe (Urbana: University of Illinois Press, 1977), 1–241. It was Suppe who introduced the phrase "the Received View" to refer to logical positivism and logical empiricism.

2. Various methods of effecting such an elimination have been proposed, including Carnap's reduction sentences, Craigian transcription, and Ramsey sentences. For a critical discussion of some of these eliminative translation projects in logical empiricism, see Israel Scheffler, *The Anatomy of Inquiry* (New York: Knopf, 1963), pt. 2, and Carl G. Hempel, "The Theoretician's Dilemma," reprinted in *Aspects of Scientific Explanation* (New York: Free Press, 1965), 173–226.

3. Historically, traditional idealism is associated most strongly with George Berkeley's view that physical objects are nothing more than collections of sensory ideas. Its phenomenalist cousin, which holds that all statements about physical objects are equivalent in meaning to claims about actual or possible sensory experience, was defended in our century by A. J. Ayer and others.

4. Consider an instrumentalist, for example, who accepts a theory *T* that includes the sentence "The temperature of a gas is directly proportional to the mean kinetic energy of the molecules composing it." This instrumentalist might take the following position: "Theory *T* is meaningful and indeed true. But do not suppose that we should take its claims purporting to refer to unobservable molecules or its theoretical terms like *kinetic energy* literally, as referring to or asserting the existence of theoretical objects and properties. To say that the terms are meaningful and the claims are true is to say that they figure instrumentally, thus-and-so, in *T*'s generation of the following set of observational consequences. . . ."

5. *The Scientific Image*, 202–3. Van Fraassen devotes his entire third chapter to a fuller explication of empirical adequacy as the claim that all actual phenomena fit within a model of the theory. A model of a theory is any structure satisfying all the axioms of the theory.

6. Compare Maxwell's argument (as we reconstructed it) with the following argument, which leans heavily on the claim that there is no nonarbitrary line to be drawn between short and tall people, since there is a continuous transition from persons who are 4'1", 4'2" . . . , to those who are 5'3", 5'4", . . . 6'5", . . . :

- 1 There are short people.
- 2 No nonarbitrary line can be drawn between short and tall people.
- 3 Therefore, we are not warranted in claiming that there are tall people.

7. Aqua regia is a mixture of concentrated nitric acid and hydrochloric acid that can dissolve the "kingly" metals, platinum and gold.

8. Note carefully that we are assuming here that the reduction of quantum mechanics to a classical theory would have to involve the strict logical derivation of quantum mechanics from that theory. For a detailed discussion of other, perhaps more appropriate notions of intertheoretic reduction, see chapter 8.

9. Or at least according to Putnam of the 1970s. Putnam later came to espouse views less sympathetic with realism.

10. Jeff Foss, "On Accepting van Fraassen's Image of Science," *Philosophy of Science* 51 (1984): 79–92. For a reply to Foss's criticism of van Fraassen, see W. Bourgeois, "On Rejecting Foss's Image of van Fraassen," *Philosophy of Science* 54 (1987): 303–8. Foss responds in his "On Saving the Phenomena and Mice: A

Reply to Bourgeois Concerning van Fraassen's Image of Science," *Philosophy of Science* 58 (1987): 278–87.

11. See John O'Leary-Hawthorne, "What Does van Fraassen's Critique of Scientific Realism Show?" *The Monist* 77 (1994): 128–45.

12. One might wonder what is "epistemological" about the five theses of CER given that truth, reference, and explanation are not usually regarded as epistemic concepts. CER would seem to be epistemological only in the weak sense that evidence is relevant to judging its truth.

13. Laudan claims that without an account of approximate truth, theses like (R1), (T1), and (T2) are "just so much mumbo jumbo" (1126). But lacking a clear account of some claim does not entail that the claim is meaningless or unintelligible. Many times in the past, scientists and philosophers have been unable to give an adequate account of claims that have later been vindicated as coherent (and sometimes true).

14. See Clyde L. Hardin and Alexander Rosenberg, "In Defense of Convergent Realism," *Philosophy of Science* 49 (1982): 604–15.

15. Mendel himself used the term *factor* to refer to the units of inheritance. The term *gene* was introduced later.

16. The following schema is what Laudan claims antirealists have long disliked about the reasoning of realists:

- 1 If T is true, then it will have true consequences.
- 2 T has true consequences.
- 3 Therefore, T is true.

This form of reasoning is clearly invalid, since the premises could be true and the conclusion false. As Laudan notes, even if the truth of a theory guarantees its having true consequences, the fact that T has true consequences cannot warrant the inference to T 's truth, since a false theory could also have true consequences. Having true consequences is necessary but not sufficient for T 's being true.

17. *Ontological* rather than *epistemological* (as Laudan had it), because the issues of truth and reference dividing realists from antirealists more closely concern what there is than how we know what there is. This is not to deny that realists and antirealists typically disagree about how epistemological considerations should constrain ontological judgments.

18. Newton-Smith develops his account in *The Rationality of Science* (London: Routledge and Kegan Paul, 1981). See especially chapter 8.

19. In his book, Newton-Smith tries to analyze relative content further by comparing the number of sentences in t_1 that are decided by T_2 with the number of sentences in t_2 that are decided by T_1 . See *The Rationality of Science*, 200–203.

20. *The Rationality of Science*, 204. Newton-Smith adds the final clause about place selection because limits of infinite sequences are often quite sensitive to the order in which their terms are selected.

21. *The Rationality of Science*, 204.

22. Strictly speaking, we cannot conclude from a failure to satisfy Newton-Smith's definition of greater verisimilitude that the theories have the same verisimilitude, but it would seem a reasonable extension of his account in cases like this to conclude that two theories of equal content have the same verisimilitude when neither has greater verisimilitude than the other.

23. Or so most philosophers believe. See the readings by Popper and Salmon in chapter 4 for a discussion of whether science can dispense with induction.

24. We encountered such a view in chapter 7, where Nancy Cartwright, in "Do the Laws of Physics State the Facts?" claims that fundamental physical laws "do not provide true descriptions of reality" but that nevertheless she has "no quarrel with theoretical entities" (867).

25. Hempel denies that narratives explain. See his distinction between the story of evolution and the theory of its underlying mechanisms in "The Thesis of Structural Identity" in chapter 6. Also, as Brown notes, Salmon no longer regards statistical-relevance relations as sufficient for an adequate explanation. See Wesley C. Salmon, *Scientific Explanation and the Causal Structure of the World* (Princeton, N.J.: Princeton University Press, 1984).

26. Notice that Hacking's account is broadly anti-Humean. According to Hume, the fundamental causal facts are observed regularities summarized in causal laws. Thus, to speak of causal relations is not to speak of connections between objects in the world, but merely to claim that some sequence of events instantiates a regularity. In denying commitment to the truth of the fundamental laws of science, while affirming the existence of causal connections among entities in the world, Hacking has rejected the Humean account. In this respect his view is similar to Nancy Cartwright's. (The Humean account and Cartwright's position are discussed in chapter 7.)

27. Putnam's theory of meaning is discussed in the commentary on chapter 7, in the section "Mellor's Defense of the Regularity Theory."

28. Ian Hacking, *Representing and Intervening*, 46. This book is an extended defense of experimental realism.

29. See note 26 above.

30. See James T. Cushing, *Quantum Mechanics: Historical Contingency and the Copenhagen Hegemony*, for a fascinating argument that quantum mechanics would have developed quite differently had realist sentiments prevailed.

31. The broadly Tarskian line, addressed in more detail in Musgrave's article than in Fine's, takes as its starting point the foundational *Convention T* according to which "snow is white" is true if and only if snow is white. For a good introduction to Tarski's theory of truth, see D. J. O'Connor, *The Correspondence Theory of Truth* (London: Hutchinson and Co., 1975).

32. Arthur Fine, "Unnatural Attitudes: Realist and Instrumentalist Attachments to Science," *Mind* 95 (1986): 149–79.

33. And even if Fine were happy with such a view, there is no obvious reason why a metaphysic of sense data should be thought inconsistent with a realist metaphysic of an objective external world or with correspondence relations holding between the world and linguistic items.

34. Assuming that we take such talk seriously and literally. We might, instead, take such talk as a philosopher's fanciful way of saying something more banal—construing, say, “The sun-as-conceived-by-Copernicus is at rest” and “The sun-as-conceived-by-Ptolemy is in motion” as (respectively) “Copernicus believed that the sun is at rest” and “Ptolemy believed that the sun is in motion.” See “The Theory-Ladenness of Observation Argument,” in the commentary on chapter 2.

35. One might object to this second point of Musgrave's, on the grounds that it presupposes something the nonrealist need not accept, namely, that that there is a moon-in-itself, in addition to the moon-as-conceived-by-us. Thorough-going relativists will find it natural to claim that there is no moon-in-itself. If there is no moon-in-itself, then Musgrave's question about its relation to the moon-as-conceived-by-us cannot even arise. Perhaps Musgrave would reply that if this route were taken, there could be no claim of contamination at all. It was, after all, Fine's original point that the information we retrieve from interaction with the moon is contaminated—information now about an interacted-with thing, not the thing-in-itself. And this requires the distinction that the nonrealist's objection disallows.

GLOSSARY

ABDUCTION (ABDUCTIVE INFERENCE) The term *abduction* was coined by C. S. Peirce to describe the inference from a puzzling phenomenon to a theory that, if true, would explain it. N. R. Hanson regarded abduction as a significant kind of logic of discovery that is neither deductive nor inductive in character. Some philosophers treat abduction as distinct from inference to the best explanation: abduction concerns a single theory and reasons for taking it seriously as an interesting candidate for further exploration; inference to the best explanation involves a comparison among several theories and reasons for believing that one of them is true or probable. Others (the majority) use the terms *abductive inference* and *inference to the best explanation* interchangeably. (See *inference to the best explanation*.)

AMBIGUITY OF FALSIFICATION The doctrine, associated with Pierre Duhem, that testable predictions can be drawn from a scientific theory only with the help of auxiliary hypotheses, theories, and assumptions. Thus, when a prediction turns out to be false, the theory might still be true, since the fault could lie with one or more of the auxiliaries. Some philosophers have espoused the much stronger thesis that no experiment or observation can falsify any theory under any circumstances. (See *underdetermination thesis*, *holism*.)

ANALYTIC STATEMENTS Modern philosophers define an analytic statement as one that has its truth (or falsity) completely determined by the meanings of the words and symbols used to express it. (Earlier definitions of analyticity by Immanuel Kant and Gottlob Frege are discussed in the commentary on chapter 3.) All tautologies (logical truths) are analytic. So, too, are statements such as "all squares have four sides" and "all mammals suckle their young." Some philosophers, notably W. V. Quine, have denied that there are any analytic statements (even in logic and mathematics) and have attacked the traditional distinction between statements that are analytic and those that are synthetic. The issue is important because analytic statements (if there are any) would be *a priori* and thus immune from refutation. (See *synthetic statements*.)

ANTIREALISM A diverse group of doctrines whose common element is their rejection of realism. In the philosophy of science, antirealism includes instrumentalism, conventionalism, logical positivism, logical empiricism, and Bas van Fraassen's constructive empiricism. Some antirealists (such as instrumentalists) deny that scientific theories that postulate unobservables should be interpreted realistically. Others (such as van Fraassen) concede that such theories should be interpreted realistically, but they deny that we should ever accept as true any theoretical claims about unobservables. (See *constructive empiricism*, *conventionalism*, *instrumentalism*.)

A POSTERIORI An epistemological concept. An *a posteriori* (or empirical) statement is one that can be known to be true (or false) only through experience: either one's own or someone else's. Apart from the mathematics and logic they contain, the empirical sciences consist of *a posteriori* statements, knowledge of which can be obtained only through observation and experiment.

A PRIORI An epistemological concept; the opposite of a posteriori. An a priori statement can be known to be true (or false) independently of experience (although some experience may be necessary to understand what the statement means). The least controversial examples of a priori statements are simple conceptual and logical truths. Examples: everything is identical with itself; all contradictions are false; all vixens are foxes.

AXIOLOGY Axiology is concerned with values. Typical axiological questions are, What things are worth pursuing? What do judgments of value mean? How can judgments about values be justified? In philosophy of science, Larry Laudan and others use the phrase *axiological level* to refer to views about the aims and goals of science.

BAYESIAN CONFIRMATION THEORY There are many different ways in which philosophers have applied Bayes's theorem to confirmation theory. The starting place for all of them is Bayes's theorem written in the form: $P(H/E \& B) = P(E/H \& B) \times P(H/B) / P(E/B)$, where E is the evidence for some hypothesis H , and B is a person's background beliefs. Objective Bayesians interpret probabilities as relative frequencies; subjective Bayesians (the majority) interpret probabilities as degrees of belief. According to the subjective Bayesians, $P(H/B)$, the prior probability of the hypothesis given one's background beliefs, will vary from individual to individual. But after acquiring the evidence, E , all rational persons should change their degree of belief in H from $P(H/B)$ to $P(H/E \& B)$. (See *likelihood*, *posterior probability*, *prior probability*, *problem of old evidence*.)

BAYES'S THEOREM In its simplest form, Bayes's theorem (also called Bayes's rule, law, or equation) asserts that $P(H/E) = P(E/H) \times P(H) / P(E)$. Although versions of Bayes's theorem were derived by the Reverend Thomas Bayes, it was first used systematically by Laplace to calculate $P(H/E)$, the conditional probability of a hypothesis, H , given evidence, E . In the twentieth century, a school of Bayesian statistics and confirmation theory has developed in which Bayes's theorem plays a central role. See *Bayesian confirmation theory*, *posterior probability*, *prior probability*.)

BRIDGE LAW Ernest Nagel and others regard the logical derivability of one theory, T , from another, T' , as a necessary condition for the reduction of T to T' . When the reduced theory, T , contains terms that are absent from the reducing theory, T' , bridge laws connecting these terms with the vocabulary of T' must be added to T' before T can be deduced from (the augmented version of) T' . The status of bridge laws is controversial. Nagel regards them as empirical hypotheses. A typical example of a bridge law is the statement connecting the temperature of a gas with the mean kinetic energy of the translational motion of its molecules, used to reduce the ideal gas law to the kinetic theory of gases. Bridge laws are sometimes called bridge principles, correspondence rules, or coordinative definitions.

CONDITION OF MEANING INVARIANCE A necessary condition for intertheoretic reduction according to Ernest Nagel and others. When one theory reduces another, it must leave all the terms of the reduced theory unchanged in meaning. The condition of meaning invariance follows directly from the requirement that the reduced theory be logically derivable from the reducing theory.

CONSISTENCY CONDITION A necessary condition for intertheoretic reduction according to Ernest Nagel and others. The reducing theory must be logically consistent with the reduced theory. Like the condition of meaning invariance, the consistency condition is a consequence of the requirement that the reduced theory be logically derivable from the reducing theory. The consistency condition has been much criticized by Paul Feyerabend on the grounds that it dogmatically protects old, well-established theories from refutation by new theories that contradict them.

CONSTRUCTIVE EMPIRICISM *Constructive empiricism* is Bas van Fraassen's term for his antirealist philosophy of science. Like logical empiricism, constructive empiricism draws a line between the theoretical and the observable. In van Fraassen's case, observable things and processes are those that can be seen by unaided human vision. Unlike the logical empiricists, van Fraassen is a realist about theories that talk about unobservables. But, van Fraassen insists, however successful these theories are in predicting and explaining phenomena, we should never accept them as true but merely as empirically adequate. (See *antirealism*, *empirical adequacy*, *logical empiricism*, *observable*.)

CONTEXT OF DISCOVERY The context of discovery includes anything that is psychologically relevant to the discovery of a theory, hypothesis, theorem, argument, or piece of evidence. (See *context of justification*.)

CONTEXT OF JUSTIFICATION The context of justification includes anything that is logically relevant (either deductively or inductively) to the epistemic justification of a theory, hypothesis, theorem, assertion about an argument, or evidential claim. The contexts of discovery and justification are not exclusive: the factors involved in the discovery of a result might play a role in justifying it, and criteria of justification often guide the search for new theories. (See *context of discovery*.)

CONTEXTUALIST ANALYSIS OF EVIDENCE The *contextualist analysis of evidence* is Helen Longino's term for the thesis, defended by T. S. Kuhn and others, that our judgments of evidential relevance and strength of evidential support for hypotheses depend on our background beliefs and assumptions. For example, what an individual scientist takes to be evidence for a hypothesis depends on the context of that scientist's beliefs. Unlike the standard view of evidence adopted by philosophers of science, on the contextualist analysis, evidence is not a set of statements but the objects, events, or states of affairs that observation statements purport to describe. (See *evidence*.)

CONTINGENCY (CONTINGENT PROPOSITION) A metaphysical concept. Contingent propositions are true in some possible worlds but not in others. True contingent propositions are true in the actual world but false in at least one other possible world. False contingent propositions are false in the actual world but true in at least one other possible world. Examples: sea water contains traces of gold; carnivorous dinosaurs prowled the streets of New York in the nineteenth century. If a proposition is not contingent, then it is necessary. (See *necessity*, *possible world*.)

CONVENTIONALISM A decision is conventional if it involves choosing from among alternatives that are equally legitimate when judged by objective criteria (such as consistency with observation and evidence); thus, either the decision is

entirely arbitrary or it rests on an appeal to factors often presumed to be subjective, such as simplicity, economy, and convenience. Radical conventionalists argue that scientific theories are really definitions (or rules of inference, pictures, conceptual schemes, paradigms) and hence neither true nor false; moderate conventionalists disagree, insisting that once the conventional elements in theories have been isolated, the remaining parts are objectively true or false. Typically, conventionalists appeal to the underdetermination of theories by evidence to bolster their doctrine. Many philosophers of science have been conventionalists in one respect or another. They include Henri Poincaré (on high-level scientific laws as definitions), Pierre Duhem (on the ambiguity of falsification of theories in physics), W. V. Quine (on the decision to retain or abandon any sentence whatever), Hans Reichenbach (on the choice of a geometry to describe physical space), Karl Popper (on the decision to accept basic statements), and T. S. Kuhn (on the decision to switch paradigms). (See *underdetermination thesis*.)

CORRESPONDENCE RULES According to logical empiricists (such as R. B. Braithwaite, Hans Reichenbach, Rudolf Carnap, Carl Hempel, and Ernest Nagel), correspondence rules give a partial interpretation of the theoretical terms of a theory by linking them with observation terms. Thus, the correspondence rules of a theory are mixed sentences that must contain (nonvacuously) at least one theoretical term and at least one observation term. (See *logical empiricism*.)

CORROBORATION *Corroboration* is a term introduced by Karl Popper to characterize the status of theories that have survived severe tests. Since Popper denies that there is any such thing as inductive confirmation, he says that theories that have survived attempts to refute them are trustworthy, not because they have been confirmed or made probable by evidence, but because they have been corroborated. Some critics have charged that either corroboration is confirmation in all but name or, if corroboration refers solely to the past success of the theory, then it can give us no rational justification for trusting that theory in the future.

COUNTERFACTUAL Philosophers of science are particularly interested in counterfactual conditionals, expressed in the subjunctive mood. For example: if my leg were made of gold, it would weigh at least two hundred pounds; if one were to slam together two pieces of uranium-235 the size of elephants, then there would be a nuclear explosion. One key difference between laws of nature and merely true but non-lawlike universal generalizations is that laws do, and nonlawlike generalizations do not, support inferences to true counterfactual conditionals. Thus, even if it were true that all dogs born at sea are cocker spaniels, it would not follow from this that if a golden retriever were to have pups at sea, they would be cocker spaniels.

COVERING LAW MODEL OF EXPLANATION The covering law model is a family of models of explanation, proposed by Carl Hempel and others, in all of which a necessary condition for a scientific explanation is that its premises contain (nonvacuously) at least one law of nature. The main examples of the covering law model are Hempel's deduction-nomological (D-N) and inductive-statistical (I-S) models of explanation. (See *deductive-nomological explanation*, *inductive-statistical explanation*.)

CRUCIAL EXPERIMENT A crucial experiment is an experiment that would conclusively falsify one of two competing theories or hypotheses, thereby estab-

lishing its rival as well confirmed or true. Pierre Duhem argued that no experiment in physics could be crucial in this sense. (See *ambiguity of falsification*.)

DEDUCTION (DEDUCTIVE VALIDITY) An argument is deductively valid if and only if its premises logically imply its conclusion. Hence, if all the premises of a valid argument are true, then the argument's conclusion must also be true. In a valid argument, it is impossible for all the premises to be true and the conclusion false.

DEDUCTIVE-NOMOLOGICAL (D-N) EXPLANATION A popular model of scientific explanation in the twentieth century, the deductive-nomological (D-N) model was given its clearest formulation and defense by Carl Hempel. D-N explanations are deductively valid arguments in which at least one premise is a scientific law. (See *epistemic conception of explanation*, *irrelevance problem*, *symmetry objection*, *thesis of structural identity*.)

DEDUCTIVE-NOMOLOGICAL MODEL OF PROBABILISTIC EXPLANATION (THE D-N-P MODEL) Peter Railton has proposed his deductive-nomological model of probabilistic explanation (the D-N-P model) as an alternative to Carl Hempel's inductive-statistical (I-S) model. Railton denies that explanations are arguments, although he insists that explanations must contain a deductive argument based on a probabilistic law. Probabilistic laws, on this view, are statements, not of frequencies, but of single-case propensities. D-N-P explanations must also contain a theoretical derivation of the relevant probabilistic law based on the indeterministic causal mechanism that brings about the event to be explained. Among the several important differences between Railton's D-N-P model and Hempel's I-S model are these: D-N-P explanations are fully objective (not relative to the state of scientific knowledge at a given time); and complete D-N-P explanations can be given of events no matter how low their probability, but they can be given only when there is genuine indeterminism at work.

DEMARICATION CRITERION A demarcation criterion, if any could be found, would distinguish genuine science from pseudoscience. Past proposals for demarcation criteria such as verifiability (the logical positivists) and falsifiability (Karl Popper) have been abandoned in light of powerful objections. Some philosophers (Imre Lakatos, Paul Thagard) think that a demarcation criterion must include historical considerations; others (Larry Laudan) think that the search for a demarcation criterion is futile.

DETERMINISM According to determinism, there is at any moment exactly one physically possible future; that is, there is exactly one set of future events that is consistent with the past and the laws of nature. A common version of determinism says that every event is causally determined by earlier events and thus, in principle, could have been predicted if we knew enough about those earlier events and the laws of nature. This way of linking determinism with predictability-in-principle has been criticized by John Earman and others on the grounds that deterministic laws do not guarantee predictability.

DISPOSITION A disposition is the power or tendency of a thing to behave in a certain way when placed in a certain kind of situation. For example, soft iron has the disposition to become magnetized when placed in a magnetic field; copper has the disposition to dissolve in hydrochloric acid; tritium (an isotope of hydro-

gen) has the disposition to undergo radioactive decay. Dispositions have been regarded with suspicion by some empiricists (such as Hume), since they are properties that a thing is supposed to have even when the disposition in question is not being manifested. Thus, the attribution of a disposition to an object goes beyond any set of evidence about how the object actually behaves. Moreover, some empiricists have questioned whether the appeal to dispositions can be genuinely explanatory.

DUHEM-QUINE THESIS It has become common to speak of the Duhem-Quine thesis as if there is one thesis to which both Pierre Duhem and W. V. Quine subscribed. This is historically doubtful. Usually the term refers to the holistic doctrine that no scientific theory can be tested in isolation, since the derivation of an observation sentence from any theory requires a host of auxiliary hypotheses, theories, and assumptions. From this, it is usually concluded that falsification is ambiguous. But some philosophers go further and conclude that the acceptance or rejection of particular scientific theories can never be justified by observational evidence. Thus, the presumed truth of the Duhem-Quine thesis is often appealed to by those who wish to emphasize the role of convention, contextual values, and social forces in the decision to accept or reject scientific theories. (See *holism*, *underdetermination thesis*.)

EMPIRICAL ADEQUACY The term *empirical adequacy* was introduced by Bas van Fraassen and plays a key role in his constructive empiricism. A scientific theory is empirically adequate when everything that it says about observables is true. As critics of van Fraassen have pointed out, to judge that a theory is empirically adequate goes far beyond claiming that the theory agrees with all the available evidence. (See *constructive empiricism*, *observable*.)

EMPIRICISM Empiricism includes a wide variety of philosophical doctrines held by philosophers such as Aristotle, Locke, Hume, Mill, and Ayer. What they share is a healthy respect for experience as the ultimate source of all human knowledge (possibly excluding logic and mathematics). Empiricists tend to be suspicious of any claims (such as those made by the theoretical sciences) that cannot be related directly to experience or observation. Some empiricists have tried, unsuccessfully, to reduce theoretical claims to claims about experience. Others (such as the logical empiricists) have tried to analyze the meaning (or cognitive significance) of theoretical assertions by tracing them to observation statements via correspondence rules. This, too, is now generally regarded as a failure. Nonetheless, empiricism in philosophy of science remains vigorous and popular.

ENTAILMENT One statement entails another when it is impossible for the second to be false when the first is true. For example, the premises of a valid argument entail its conclusion.

EPISTEMIC CONCEPTION OF EXPLANATION Wesley Salmon gives Carl Hempel's covering law model as a classic example of the epistemic conception of explanation. On this conception, explanations are a type of argument or inference, and to explain something is to show why that thing was to be expected given the information in the premises of the explanatory argument. It is characteristic of the epistemic conception that explanation and prediction are seen as formally identical and that explanations can be given for events only if their occurrence is certain or highly probable. (See *ontic conception of explanation*.)

EPISTEMIC REGULARITY THEORY OF LAWS A version of the Humean, regularity theory of laws. According to A. J. Ayer and others, laws of nature are true, universal generalizations that have additional features of an epistemic character. These features include our willingness to use the generalization to make predictions and the role that the generalization plays in an organized system of science. (See *regularity theory of laws*.)

EPISTEMOLOGY Epistemology is the theory of knowledge. Epistemologists concern themselves with concepts such as knowledge, belief, epistemic justification, confirmation, and explanation. Epistemic problems include refuting skepticism (the denial that knowledge, either in general or in a particular area, is possible) by explaining why some beliefs (and not others) are justified.

EVIDENCE For most philosophers of science, evidence is an observation statement (or a set of such statements) reporting the outcome of an observation or experiment that either confirms or disconfirms a theory. Positive evidence confirms; negative evidence disconfirms. (See *contextualist analysis of evidence*.)

EXPLANANDUM A Latin word meaning "that which is to be explained." The explanandum is usually taken to be a sentence describing the thing (whether event, law, or theory) for which an explanation is sought. (See *explanans*.)

EXPLANANS A Latin word meaning "that which explains." The explanans is the thing, usually regarded as a set of sentences satisfying certain conditions, that constitutes an explanation of the explanandum. (See *explanandum*.)

EXPLANATIONISM Explanationism is a thesis about confirmation. Explanationists insist that only the explanation of previously known facts has the power to confirm a theory, thus denying that a theory receives any confirmation from novel predictions. (See *historical thesis of evidence*, *novel prediction*, *predictionism*.)

EXTENSION The extension of a term (predicate, concept) is the thing or set of things that the term picks out and to which the term is correctly applied. Thus, the extension of *yellow* is the set of yellow things. The extension of a proper name is its bearer. (See *intension*.)

FALSIFICATIONISM A set of methodological doctrines championed by Karl Popper comprising a demarcation criterion and a fallibilist theory of scientific rationality and progress. According to Popper, a theory is scientific if and only it is falsifiable. No theory can be proven to be true or even probable. Scientific rationality consists in testing theories severely in order to refute them. Progress occurs when falsified theories are replaced by as-yet-unfalsified theories that are more testable than those they replace. A more sophisticated version of falsificationism has been advocated by Imre Lakatos. (See *corroboration*, *scientific research programme*.)

FUNCTIONAL LAWS Functional laws assert a universal relation between two or more variables in the form of a mathematical equation or inequality. Typical examples are Hooke's law, Snell's law, and Maxwell's field equations.

HIERARCHICAL MODEL OF SCIENTIFIC RATIONALITY A model of scientific rationality, criticized by Larry Laudan, according to which decisions at one level are justified (if at all) by rules and criteria at the next highest level. Inevitably, on this model disagreements about the aims and goals of science at the highest,

axiological level cannot be resolved rationally. (See *axiology*, *reticulation model of scientific rationality*.)

HISTORICAL THESIS OF EVIDENCE The historical thesis of evidence maintains that whether or not a piece of evidence confirms a theory depends on the time that the evidence is known relative to the time at which the theory is proposed. Both explanationism and predictionism entail the historical thesis of evidence. This thesis has been criticized by Peter Achinstein and Laura Snyder, and is rejected by Bayesians as well as by all those (such as Rudolf Carnap) who regard confirmation as a formal or logical relation between propositions. (See *explanationism*, *predictionism*.)

HOLISM (WHOLISM) A family of doctrines whose common core is the notion that an individual element in a complex whole has the properties it does only insofar as it stands in certain relations to other, similar elements. For example, W. V. Quine is a semantic holist, holding that the meaning of an individual sentence depends on how it is related to other sentences in the same language. Similarly, in epistemology, proponents of coherence theories of justification insist that a particular belief is justified only if it coheres with a large number of similar beliefs. In the philosophy of science, the term *holism* usually refers to Pierre Duhem's doctrine that no individual physical theory (and, more generally, no individual scientific theory) implies any observation statement. Single theories make testable predictions only when they are conjoined with other theories, auxiliary hypotheses, and background assumptions.

HOMOGENEOUS REDUCTION Ernest Nagel calls the reduction of one theory to another homogeneous when the vocabulary of the reduced theory is included in (or can be defined in terms of) the vocabulary of the reducing theory, thus permitting the logical derivation of the one theory from the other. An alleged example of a homogeneous reduction is the derivation of Galileo's law of falling bodies from Newton's gravitational theory. (See *inhomogeneous reduction*.)

HYPOTHESIS Some authors distinguish between hypotheses and theories in the following way: a hypothesis is a scientific claim that is put forward as a plausible conjecture before there is enough evidence to warrant either its acceptance or its rejection; a theory is a well-established scientific claim that is accepted as true or well-confirmed. Hypotheses, if successful, become theories. Other authors (such as the authors of the present book) use the terms *hypothesis* and *theory* interchangeably. (See *theory*.)

HYPOTHETICO-DEDUCTIVE (H-D) MODEL In its simplest and most common form, the hypothetico-deductive (H-D) model of science denies that there is any logic of discovery and affirms we are justified in accepting or rejecting theories only after they have been tested. Testing a theory consists of deriving from it consequences that can then be compared with observations and experimental results. If the results agree with the predictions, the theory is inductively confirmed; if the results disagree, the theory is disconfirmed or refuted.

IDENTITY So-called identical twins are qualitatively identical (having, let us assume, all the same qualitative properties) but not numerically identical (since they are two distinct individuals). When philosophers of science talk about identity,

they usually mean numerical identity. For example, flashes of lightning are identical with discharges of atmospheric electricity, genes are identical with segments of DNA, and water is identical with molecules of hydrogen oxide. If an identity statement is true, then the phrases used to express it (such as "a flash of lightning" and "a discharge of atmospheric electricity") have the same extension—they both refer to the same thing. This is not the same as claiming, nor does it entail, that the phrases mean the same thing or that anyone using the terms correctly must be aware that the terms are coextensive. On the rigid designator theory proposed by Saul Kripke and Hilary Putnam, statements of identity are metaphysically necessary but the discovery that they are true can (and, in the empirical sciences, usually does) depend on observation and experiment. (See *extension*, *necessity*, *rigid designator*.)

INCOMMENSURABILITY A variety of doctrines fly under the banner of incommensurability. Prominent among them is the claim of Paul Feyerabend and T. S. Kuhn that the meaning of a theoretical term depends on the theory in which it occurs in such a way that it is impossible for rival theories to share meanings. Another variety of incommensurability, defended by Kuhn, is the alleged incommensurability of standards for choosing among theories (what Kuhn calls *values*) associated with rival paradigms.

INDUCTION Induction can be construed broadly or narrowly. Broadly, it is any deductively invalid inference in which the premises make the conclusion probable; narrowly, it is any inference that is an instance of argument forms such as induction by simple enumeration, argument by analogy, statistical syllogism, and induction to a particular. The broad characterization includes as inductive, and the narrow may exclude, argument types such as inference to the best explanation.

INDUCTIVE-STATISTICAL (I-S) EXPLANATION An influential model of explanation, advocated by Carl Hempel, in which the explanans renders the explanandum highly probable but does not deductively entail it. The explanans of an I-S explanation has to include a statistical law, and the form of the argument is a statistical syllogism.

INDUCTIVISM A hopelessly naive model of science according to which scientific knowledge accumulates by using inductive generalization to infer laws and theories from observational and experimental facts. According to inductivism, induction serves as both the logic of discovery and the logic of justification. Inductivism was decisively refuted by Pierre Duhem and Karl Popper. Although everyone agrees that inductivism is false, many philosophers of science insist (contrary to Popper) that the inductive confirmation of theories by the verification of their logical consequences is an essential part of scientific reasoning. (See *hypothetico-deductive model*.)

INFERENCE TO THE BEST EXPLANATION A form of argument often relied on by scientific realists in which the fact that a particular theory explains some body of evidence better than any of its rivals is supposed to make the theory probable or reasonable to believe. Usually, inference to the best explanation is regarded as distinct from inductive reasoning. Some of its proponents (such as Gilbert Harman and Peter Lipton) have argued that it underlies most of the reasoning traditionally regarded as inductive. (See *abduction*.)

INHOMOGENEOUS REDUCTION According to Ernest Nagel, inhomogeneous reductions are those in which the vocabulary of the reduced theory is neither included in nor can be defined in terms of the vocabulary of the reducing theory. Nagel argues that the one theory can be logically derived from the other (and thus be reduced by it) only when the reducing theory is supplemented with appropriate bridge laws. An alleged example of an inhomogeneous reduction is the derivation of the ideal gas laws from the kinetic theory of gases. (See *bridge law*, *homogeneous reduction*, *reductionism*.)

INSTRUMENTALISM Strictly speaking, instrumentalism is the doctrine that theories are merely instruments, tools for the prediction and convenient summary of data. As such, theories are not statements that are either true or false; they are tools that are more or less useful. But because one has to use the machinery of logic in order to draw predictions from theories, it is difficult to deny that theories have truth values. Thus, instrumentalism has come to be used as a general term for antirealism. Most modern instrumentalists concede that theories have truth values but deny that every aspect of them should be interpreted realistically or that reasons to accept a theory as scientifically valuable are reasons to accept the theory as literally true. In this sense, T. S. Kuhn, who locates the value of scientific theories in their ability to solve puzzles, is an instrumentalist. Theories may have truth values, but their truth or falsity is irrelevant to our understanding of science. (See *antirealism*, *constructive empiricism*, *conventionalism*, *scientific realism*.)

INTENSION A semantic notion; the intension of a general term or concept, A, is its meaning, connotation, or sense. Some authors define the intension of A as all the general terms or concepts B such that "All As are B" is a necessary truth. Others insist on a narrower characterization that restricts the Bs to properties that belong to the explicit definition of A. Still others (such as W. V. Quine) reject the whole notion of intension and meaning as a seductive myth. (See *extension*, *sense*.)

INTENTIONALITY The intentionality of beliefs and other mental states is the property that enables them to be about or to represent states of affairs. The intentionality of beliefs (hopes, fears, wishes) is usually indicated by a "that" clause followed by a proposition. If Sarah believes (hopes, fears, wishes) that the dodo is not extinct, then the intentional content of her belief (hope, fear, wish) is the nonactual state of affairs of the dodo's not being extinct.

IRRELEVANCE PROBLEM A general term for a variety of alleged counterexamples to Carl Hempel's deductive-nomological (D-N) and inductive-statistical (I-S) models of explanation. These cases satisfy Hempel's conditions for an adequate explanation but in each of them the explanans is explanatorily irrelevant to the explanandum.

LAWS OF NATURE Scientists discover laws of nature but philosophers disagree about what kinds of thing laws of nature are. The main competitors are the regularity theory, the epistemic regularity theory, the necessitarian theory, and the universals theory of laws (all defined elsewhere in this glossary). Laws of nature are not restricted to the things that modern scientists call laws; they include all (true) fundamental equations and inequalities, whether or not they are called laws and regardless of whether they have been or ever will be discovered.

LIKELIHOOD The likelihood of a theory, T, relative to some evidence, E, and background information, B, is $P(E/T\&B)$, the probability of the evidence given

the theory and the background information. When *T* entails *E*, the likelihood of *T* (relative to *E*) is 1.

LOGICAL EMPIRICISM Some authors use the labels *logical empiricism* and *logical positivism* interchangeably. Others (such as the authors of this book) reserve the term *logical empiricism* for the views of Rudolf Carnap, Carl Hempel, Ernest Nagel, Hans Reichenbach, and others after the verifiability principle was abandoned. Among the doctrines of logical empiricism are the theory-observation distinction, the view that theoretical terms are “partially interpreted” (through their role in the derivation of observation statements), Hempel’s deductive-nomological and inductive-statistical models of explanation, and Nagel’s model of reduction. (See *logical positivism*, *verifiability principle of meaning*.)

LOGICAL POSITIVISM Logical positivism is the name for the set of doctrines advocated by the members of the Vienna Circle from about 1920 to 1936 (when Moritz Schlick, their leader, was assassinated). Prominent members of this group were Rudolf Carnap, Herbert Feigl, Hans Hahn, Otto Neurath, and Friedrich Waismann. Their approach to philosophy relied heavily on the verifiability criterion of meaning and is well illustrated in A. J. Ayer’s *Language, Truth and Logic*. (See *logical empiricism*, *verifiability principle of meaning*.)

LOGICISM The thesis, advocated by Gottlob Frege and Bertrand Russell, that mathematics (or, at least, arithmetic) can be reduced to logic by appropriate definitions of terms such as *number* and *is a successor of*. The logicist program foundered, in part because arithmetic requires set theory and the axioms of set theory are not truths of logic.

MEANING INVARIANCE, CONDITION OF Ernest Nagel imposes meaning invariance as a condition for intertheoretic reduction: when one theory reduces another it must leave the meanings of the terms in the reduced theory unchanged.

METAPHYSICS The branch of philosophy that attempts to discover necessary truths about the fundamental constituents of reality (or, more cautiously, about things that, if they exist, would be basic features of the world). Metaphysics includes inquiries into the nature of time, space, matter, causation, minds, freedom, and God and the relations between them. It is widely acknowledged that there is no rigid line separating metaphysics from science (especially physics).

METHODOLOGY In the philosophy of science, *methodology* means the *theory of scientific method*. For example, methodological principles are the rules that are supposed to govern rational choices among scientific theories. More generally, methodology includes criteria for assessing theories, weighing evidence, and evaluating scientific change.

MODUS PONENS The traditional name for any valid argument of the form, “If *P* then *Q*; *P*; therefore, *Q*.” (See *deduction*.)

MODUS TOLLENS The traditional name for any valid argument of the form, “If *P* then *Q*; not-*Q*; therefore, not-*P*.” (See *deduction*.)

NATURAL KINDS Speaking very generally, natural kinds are classes of naturally occurring things that form the proper object of scientific theorizing and laws. They include chemical elements, fundamental particles, and biological species. Decisions about which things belong to natural kinds often depend on our best sci-

entific theories. Oxygen, diamonds, human beings, and neutron stars are natural kinds; phlogiston, valuable gems, U.S. residents, and things that shine in the night sky are not.

NATURAL ONTOLOGICAL ATTITUDE (NOA) The natural ontological attitude (acronym, NOA, pronounced “Noah”) is advocated by Arthur Fine as a solution to the debate between realism and antirealism in the philosophy of science. According to Fine, NOA is the minimal, neutral “core position” that is shared by both realism and antirealism but is itself neither realist nor antirealist. The NOAer (pronounced “knower”) accepts the confirmed theories and results of science (as true) in the same way that we accept everyday truths on the evidence of our senses but refuses to adopt any metaphysical theory about the nature of truth (such as truth as correspondence with reality).

NECESSARY CONDITION X is a necessary condition for Y if and only if one cannot have Y without X. Being a reptile is a necessary condition for being an alligator; the presence of oxygen is a necessary condition for the rusting of iron.

NECESSITARIAN THEORY OF LAWS The necessitarian theory of laws insists that there must be a real, objective, relation of necessity (*nom*ic necessity) between events, objects or properties, if a statement about them can be a law of nature. Necessitarians differ among themselves about whether laws of nature are universal generalizations about particulars or singular statements of fact about universals. But all necessitarians agree on the necessity of laws of nature and reject the regularity account. (See *epistemic regularity theory of laws*, *regularity theory of laws*.)

NECESSITY A much-contested metaphysical notion. Propositions are either logically, metaphysically, or physically necessary (where these categories are neither exclusive nor exhaustive). Metaphysically necessary propositions (a category that includes, but is not confined to, all those propositions that are logically necessary) cannot be false; they are true in all possible worlds. Physically necessary propositions are those whose truth is guaranteed by the laws of nature. Propositions that are neither metaphysically nor logically necessary are said to be contingent. (See *contingency*.)

NONINSTANTIAL LAWS Noninstantial laws are laws that have no instances. Newton’s first law (the law of rectilinear inertia) is often cited as an example because, it is claimed, no actual body is ever completely free of a net external force. Nonetheless, Newton’s law is true: if there were such a body, it would not accelerate. Likewise, it is a lawlike fact that all pieces of plutonium with a mass greater than 1 million kilograms are excellent conductors of electricity even though the universe never has and never will contain such objects. (See *vacuous laws*.)

NORMAL SCIENCE A term introduced by T. S. Kuhn to describe science the way it is most of the time, when no scientific revolutions are occurring. During normal science, Kuhn claims that scientists in a given field work under the aegis of a single paradigm that they all accept. (See *paradigm*.)

NOVEL PREDICTION The prediction of a result by a theory can be novel in many different senses. One straightforward sense, temporal novelty, is that the result predicted was not known by anyone at the time the theory was proposed. Another sense, epistemic novelty, is that the result was not widely known and, in particular, was unknown to the person proposing the theory. Two other senses

deemed important by some philosophers of science are design-novelty and use-novelty. A result is design-novel (or heuristically novel) when the theory yielding it was not deliberately constructed so as to yield that result (among others). A result is use-novel when it was not built into the new theory by, for example, using it to fix the value of a parameter. Philosophers of science disagree about whether novelty (in one or more of these senses) is a necessary condition for confirmation or, even if it is not necessary for confirmation, whether novelty enhances the confirming power of a prediction. (See *predictionism*.)

OBSERVABLE To paraphrase John Stuart Mill, something is observable when it can be observed. Opinions differ as to what this amounts to and what its epistemic significance is. The logical positivists and empiricists defined *observation terms* as those terms whose correct application can be rapidly decided by any person of normal perceptual powers without any special scientific training. Bas van Fraassen defines *observable objects* as those that can be seen by the unaided vision of a normal human being. The intended contrast is between those terms and things that are observable and those that are theoretical. For empiricists, theoretical claims ultimately rest on and are tested by observation reports confined to observables. (See *constructive empiricism*, *theory-ladenness*.)

ONTIC CONCEPTION OF EXPLANATION Wesley Salmon contrasts the ontic conception of explanation (which he endorses) with the epistemic conception. Unlike the epistemic conception, the ontic conception denies that explanations are arguments (although they may involve arguments) and rejects the requirement that the explanandum be predicted with high probability. On the ontic view, something is an explanation only if it correctly describes the underlying cause or causal mechanism that brought about the explanandum event. A good example of an ontic approach to explanation is Peter Railton's deductive-nomological model of probabilistic explanation (D-N-P model). (See *deductive-nomological model of probabilistic explanation*, *epistemic conception of explanation*.)

ONTOLOGY Either the part of metaphysics concerned with the nature of existence, or, as in the philosophy of science, the entities (things, processes, properties) postulated by a particular scientific theory or conceptual scheme.

OPERATIONALISM (OPERATIONISM) The doctrine propounded by the physicist Percy W. Bridgman in his book, *The Logic of Modern Physics* (1927). Impressed by the success of Einstein's relativity theory, especially Einstein's analysis of the concept of simultaneity, Bridgman proposed that all scientific concepts and terms are synonymous with the operations and procedures used to measure their values. But as critics such as Carl Hempel have pointed out, very few scientific concepts can be defined completely in terms of a particular set of procedures and, in any case, Bridgman's proposal would lead to the unnecessary proliferation of scientific concepts if each new measuring procedure were to be regarded as defining a new concept.

PARADIGM A multiply ambiguous term that lies at the heart of T. S. Kuhn's philosophy of science. In most of its uses, it means either *exemplar* or *disciplinary matrix*. An exemplar is an important scientific theory or piece of research that serves as a model for further inquiry. Disciplinary matrices contain exemplars as one of their elements. They also include heuristic models, ontological and meta-physical assumptions, and methodological principles. In short, paradigms (in the

sense of disciplinary matrices) are “super theories” that underlie and guide an entire tradition of scientific research and theorizing.

PHENOMENALISM Proposed in the eighteenth century by Bishop Berkeley and in the twentieth by the logical positivists, phenomenalism is the doctrine that all meaningful talk about physical objects is reducible to talk about actual or possible perceptual experience. (This is the linguistic version of phenomenalism; its ontological version asserts that physical objects are nothing more than sets of actual and possible sense experiences.) John Stuart Mill, Ernst Mach, A. J. Ayer, Bertrand Russell and many other philosophers have been phenomenologists at some point in their careers. It is now generally acknowledged that it is impossible to give the translations required by the linguistic version of phenomenalism.

POSITIVISM An extreme form of empiricism advocated by the French philosopher and sociologist Auguste Comte (1798–1857). Comte denied that it is possible to know anything about unobservables (into which category he placed the underlying causes of phenomena), and he insisted that the sole aim of science is prediction, not explanation. Comte also believed that each branch of science has its own special laws and methods that cannot be reduced to those of other branches. Generally regarded as the founder of sociology, Comte’s empiricism and his hostility towards metaphysics were an important influence on logical positivism. (See *logical positivism*.)

POSSIBILITY (LOGICAL, PHYSICAL) A statement or a state of affairs is logically possible if its denial is not logically necessary. Something is physically possible if, relative to a given set of circumstances, the laws of nature do not forbid it. What is physically possible in one set of circumstances might be impossible in another. Many logically possible things are physically impossible, but everything that is physically possible is logically possible. (See *necessity*.)

POSSIBLE WORLD The actual world (*this* world) is a possible world. So, too, is any world whose complete description does not contain any logically impossible proposition. (For every proposition, a complete description of a possible world contains either that proposition or its denial.) (See *contingency*, *possibility*.)

POSTERIOR PROBABILITY The posterior probability of a theory, *T*, relative to some evidence, *E*, and background information, *B*, is $P(T/E\&B)$, the probability of the theory given the evidence and the background information. (See *Bayesian confirmation theory*, *prior probability*.)

PREDICATE In the sentence, “Sodium is monovalent,” *sodium* is the subject term and *monovalent* is the predicate term. Similarly, “Socrates is married to Xanthippe” predicates of Socrates the property of being married to Xanthippe. Philosophers disagree about whether every predicate expresses a genuine property. Some authors use the term *predicate* to include not only monadic predicates, but also relations (two-place predicates and higher).

PREDICTIONISM According to predictionism, when a result is derived from a theory, the result confirms the theory only if it is a novel prediction; a theory is not confirmed by “old evidence,” no matter how much of it the theory is able to explain. A less extreme form of predictionism is accommodationism—the view that while the explanation or accommodation of old evidence can have some confirm-

atory value, the prediction of new evidence always has more. (See *explanationism*, *novel prediction*, *problem of old evidence*.)

PRIOR PROBABILITY The prior probability of a theory is sometimes taken to be the theory's probability when it was first proposed, before any evidence for or against it has been considered. But often the prior probability of a theory means the theory's probability prior to the consideration of a particular piece of evidence. Thus, in the first sense, a theory has only one prior probability, but in the second sense, prior probabilities change as evidence accumulates. (See *Bayesian confirmation theory*.)

PROBLEM OF OLD EVIDENCE The problem of old evidence, for Bayesian confirmation theorists, is that of explaining why we should not set equal to 1 the probability of any piece of evidence that we already know to be true. The problem is that, on the Bayesian analysis of confirmation, this would prevent that evidence from confirming any theory. (See *Bayesian confirmation theory*.)

REALISM One can be a realist about many different kinds of thing: numbers, possible worlds, universals, minds, physical objects, quarks, fields, and so on. To call a philosopher a realist requires a specification of what the philosopher is a realist about. Usually, there is an intended contrast with those who deny that the entities in question are real. In the philosophy of science, realists are aligned against instrumentalists, phenomenologists, conventionalists, fictionalists, and others of that ilk. (See *scientific realism*.)

REDUCTIO AD ABSURDUM (REDUCTIO ARGUMENT) A form of valid deductive reasoning in which a claim is refuted by deducing from it a false, contradictory, or "absurd" conclusion. (If the deduction requires additional premises, those statements must be known to be true for the reductio reasoning to be sound.) In mathematics and logic, reductio reasoning is often called the *method of indirect proof*. (See *deduction*.)

REDUCTIONISM Philosophers of science use the term *reductionism* in at least three different senses. Either it refers to the research strategy of trying to understand complex systems by studying their components and the relations among them; or it refers to the claim that a particular scientific theory (such as classical genetics) has been reduced to another theory (such as molecular biology) according to a particular model of intertheoretic reduction (usually the standard model advocated by Ernest Nagel and others); or it refers to the more general claim that scientific knowledge grows primarily by theory reduction (again, as judged by a model of intertheoretic reduction such as Nagel's). The first two senses are quite distinct. The success or failure of reductionism (as a research strategy) has no direct bearing on whether intertheoretic reduction has occurred or is likely to occur: reductionism may be a fruitful research strategy without resulting in theory reduction, and theory reductions may occur without being the result of a reductionistic research strategy. As discussed in chapter 8, several philosophers of biology (notably David Hull and Philip Kitcher) have opposed reductionism in the second sense by arguing that classical genetics cannot be deduced from molecular biology (as Nagel's model of reduction requires), and many philosophers of science (especially Paul Feyerabend and T. S. Kuhn) have criticized the claim that scientific knowledge grows primarily by theory reduction. (See *homogeneous reduction*, *inhomogeneous reduction*.)

REGULARITY THEORY OF CAUSATION A theory of causation, also called the constant conjunction theory, associated most notably with David Hume. In its simplest form, the theory asserts that an event of some kind *F* causes an event of some kind *G* if and only if *F*-events are, as a matter of empirical fact, always followed by *G*-events. This view denies that there is any kind of causal necessity or power in objects in virtue of which one event or object produces another. The regularity theory of causation may thus be regarded as a conjunction of two claims: (1) causal connections are a species of lawlike connections, and (2) laws of nature merely express regularities of behavior. (See *regularity theory of laws*.)

REGULARITY THEORY OF LAWS A theory of laws of nature, associated with David Hume, which holds that a law of nature is merely a contingently true universal generalization describing how all objects of a certain kind always behave. This view denies that laws express any kind of causal necessity. The regularity theory of laws is part of the regularity theory of causation. An important problem for the regularity theory of laws is its apparent inability to distinguish between genuine laws of nature (e.g., all bodies free from any net external force remain at rest or in uniform motion in a straight line) and merely accidentally true generalizations (e.g., all U.S. presidents are men). (See *counterfactual, epistemic regularity theory of laws, regularity theory of causation, universal generalization*.)

RELATIVISM Moral relativists deny that there are any objective ethical standards for evaluating human conduct independently of particular societies, their beliefs, and institutions. In the philosophy of science, epistemological relativists deny that there are any objective methodological standards for evaluating theories independently of particular scientific research traditions and their associated belief systems. Probably the best known epistemological relativist in the philosophy of science is T. S. Kuhn, who insists that all decisions about theories are relative to particular paradigms.

RELEVANCE CRITERION OF CONFIRMATION A central feature of the Bayesian approach to confirmation, the relevance criterion holds that a piece of evidence, *E*, confirms a hypothesis, *H*, if and only if *E* raises the probability of *H*. Or, in other words, *E* confirms *H* if and only if the posterior probability of *H* on *E* is greater than the prior probability of *H*. (See *posterior probability, prior probability*.)

RETICULATIONAL MODEL OF SCIENTIFIC RATIONALITY A model of scientific rationality proposed by Larry Laudan. The distinctive features of the reticulatory model are its nonlinear conception of justification and its piecemeal approach to the evaluation of values and rules at every level. (See *hierarchical model of scientific rationality*.)

RIGID DESIGNATOR An expression is a rigid designator if and only if it has the same reference in every possible world in which it has any reference at all. According to Saul Kripke and Hilary Putnam, proper names (e.g., Marie Curie) and natural-kind terms (e.g., radium) are rigid designators, but definite descriptions (e.g., the discoverer of radium) are not.

SATISFACTION CRITERION OF CONFIRMATION The satisfaction criterion is a qualitative theory of confirmation proposed by Carl Hempel. The technical details are complicated, but the basic idea is simple: a hypothesis is confirmed by

an observation report if the individuals mentioned in the report satisfy the hypothesis. Satisfaction includes, but is not limited to, positive instances. A hypothesis is satisfied by (and hence confirmed by) an observation report if the report entails a description of the world that would be true if the hypothesis were true and if the world consisted only of the individuals that are mentioned, essentially, in the report. Thus, for example, the hypothesis that all ravens are black is confirmed by the report that Adam is a black raven. Now consider the observation report that Beth is not a raven (without saying anything about her color). This report entails that either Beth is not a raven or Beth is black. And that would be a correct description of the world if Beth were the only individual in the world and if it were true that all ravens are black. (This is because the universal generalization "all ravens are black" is logically equivalent to "for all things, either that thing is not raven or that thing is black.") Thus, according to Hempel's satisfaction criterion, the hypothesis that all ravens are black is confirmed (to some positive degree, however small) not only by the observation of black ravens but also by the observation of nonravens and by the observation of nonblack things (and by the observation of things that are both nonblack and nonravens). Hempel's satisfaction criterion has several controversial features: it requires adopting the regularity theory that laws are nothing more than true universal generalizations, and its application is restricted to hypotheses that can be expressed using the observation terms that appear in observation reports. Many philosophers also find highly counterintuitive Hempel's judgment that the observation of a white shoe, say, can confirm that all ravens are black. (See *regularity theory of laws*.)

SCIENTIFIC REALISM Scientific realism has several dimensions: metaphysical, epistemological, and methodological. While there is no single, monolithic version of scientific realism that all scientific realists accept, scientific realism is generally taken to be the doctrine that the world studied by science exists and has the properties it does independently of our beliefs, perceptions, and theorizing; that the aim of science is to describe and explain that world, including those many aspects of it that are not directly observable; that, other things being equal, scientific theories are to be interpreted literally; that to accept a theory is to believe that what it says about the world is true, and that by continually replacing current scientific theories with better ones, science makes objective progress and its theories get closer to the truth.

SCIENTIFIC RESEARCH PROGRAMME A term coined by Imre Lakatos, usually spelt in the English fashion. According to Lakatos, scientific research programmes have three components: a hard core, a protective belt, and a positive heuristic.

SCIENTIFIC REVOLUTION Scientific revolutions are important scientific advances, such as those associated with Copernicus, Newton, Darwin, and Einstein, in which fundamental concepts and theories were replaced by radical new ones, dramatically changing the course of science. People talked about scientific revolutions long before T. S. Kuhn, but it is Kuhn's views about the nature of those revolutions that has made the term popular. According to Kuhn, scientific revolutions are noncumulative and the new paradigm is incommensurable with the one it replaces. (See *incommensurability*, *normal science*, *paradigm*.)

SEMANTICS The study of the meaning of signs and their relation to what they signify. Semanticists are especially interested in language. They offer theories about how words and sentences mean what they do and what makes sentences true. Meaning, truth, sense, reference, synonymy, and analyticity are all semantic concepts. (See *syntax*.)

SENSE Following Gottlob Frege, sense is usually contrasted with reference. For example, the terms *Epsom salts* and *magnesium sulfate* differ as to sense but refer to the same chemical compound. Frege thought that all referring expressions have a sense and that it is in virtue of this sense that they refer. More recent philosophers (such as Saul Kripke) disagree, holding that some proper names refer even though they lack a sense. Sense is often referred to as meaning or intension. (See *extension*, *intension*, *rigid designator*.)

SENSE DATA *Sense data*, the plural of *sense datum* (from the Latin *datum*, meaning “given”), are what we are immediately aware of when we perceive something—or think we perceive something—through our senses (e.g., colored shapes, smells, noises). David Hume called them *impressions*. Sense data are supposed to be the raw content of experience, before it has been conceptualized or described in words. Although in the past empiricists regarded sense data as the basis for all human knowledge, the psychological reality of sense data is now widely doubted, even by empiricists.

SOUNDNESS (OF AN ARGUMENT) An argument is sound if and only if it is valid and all its premises are true. Necessarily, therefore, all sound arguments must have true conclusions. The problem is to figure out which arguments are sound.

STRONG PROGRAMME The strong programme is a movement in the sociology of science that professes to make the study of science “scientific” by tracing the psychological and sociological causes of scientific beliefs and decisions, especially decisions to accept or reject theories. Particularly controversial among philosophers of science is the strong programme’s insistence that all scientific beliefs, whether true or false, rational or irrational, should be explained in the same sort of way in terms of social and cultural factors. The phrase “strong programme” was coined by David Bloor, one of the founders of the Science Studies Unit at Edinburgh University, and many of its proponents are sociologists and historians of science working in the unit.

SUFFICIENT CONDITION If *X* is a sufficient condition for *Y*, then anything that is *X* must also be *Y*. Having atomic number 88 is sufficient for an atom to be radium; a sufficient condition for the production of alpha particles is the presence of radium.

SYMMETRY OBJECTION A blanket term for a variety of alleged counterexamples to the thesis of structural identity. (See *thesis of structural identity*.)

SYMMETRY THESIS See *thesis of structural identity*.

SYNTAX The syntax of a language is its grammar. In logic, the syntax of a formal system includes rules for constructing well-formed formulas and rules for the derivation of theorems from the axioms of the system. Unlike semantics, syntax is concerned solely with rules for arranging signs, especially words and sentences, independent of what they signify or mean. (See *semantics*.)

SYNTHETIC STATEMENTS Synthetic statements are all the ones that are not analytic. On modern accounts of analyticity, this entails that the truth (or falsity) of synthetic statements is not solely a consequence of the words and symbols used to express them. (See *analytic statements*.)

TAUTOLOGY Any statement that is true solely in virtue of its logical form, e.g., if tigers are herbivorous, then tigers are herbivorous; either xenon combines with phosphorus or it does not. Sometimes the term *tautology* is used more broadly to refer to analytic statements that are true by definition, e.g., no herbivores eat flesh. (See *analytic statements*.)

THEORY Opinions differ as to what a scientific theory is and how it should be analyzed. The traditional view, and still the view of most philosophers of science in the twentieth century, is that a theory is a set of statements or propositions. For example, Newton's theory of gravity consists of a statement of Newton's law of gravitational attraction together with a few other statements essential to Newton's theory. A newer view, the semantic view advocated by Bas van Fraassen, Ronald Giere, and others, is that a theory is a definition or a set of models. In any case, to call something a theory is not to denigrate it or to imply any unfavorable contrast with facts. A theory becomes a fact (or, at least, is believed to be a fact) when it is well confirmed and established. (See *hypothesis*.)

THEORY-LADENNESS According to some philosophers of science (such as Paul Feyerabend, T. S. Kuhn, and N. R. Hanson), observations are theory laden. What exactly this means is controversial, and whether it is true and what it implies has been hotly debated. One fairly uncontroversial reading of the theory-ladenness thesis is that observation reports, if they are to be evidentially relevant to a theory, must be expressed in the vocabulary of that theory rather than in some theory-neutral observation language. Some philosophers have inferred from this that because observation is theory laden it cannot be appealed to as a neutral ground on which to judge among competing theories. Less tendentiously, it is sometimes claimed, under the rubric of the theory-ladenness-of-observation thesis, that observation and experiment alone cannot rationally compel scientists to accept and reject theories.

THESIS OF STRUCTURAL IDENTITY The thesis of structural identity (sometimes called the symmetry thesis) is the claim that prediction and explanation are formally identical, consisting in the same type of argument. The thesis comprises two claims: every adequate explanation is potentially a prediction, and every adequate prediction is potentially an explanation. Carl Hempel has been the foremost defender of the thesis of structural identity, since it is a corollary of his deductive-nomological (D-N) model of explanation, but he admits that there is some doubt about whether all adequate predictions are potentially explanations.

UNDERDETERMINATION THESIS The thesis, associated with Pierre Duhem and W. V. Quine, that neither the truth nor the falsity of any scientific theory is determined by evidence. From this it is often concluded (by T. S. Kuhn, David Bloor, and others) that the decision to accept or reject scientific theories must depend on sociological factors. Larry Laudan usefully distinguishes between deductive and ampliative versions of the underdetermination thesis, arguing that, while the former is harmlessly true, the latter is perniciously false. (See *ambiguity of falsification*, *holism*, *strong programme*.)

UNIVERSAL GENERALIZATION A universal generalization is any statement of the form "All As are B." In first-order predicate logic, this is written as $(x)(Ax \supset Bx)$, where " \supset " stands for the material conditional of truth-functional logic, and the formula is read as "for all things x , if x has property A then x has property B."

UNIVERSALS Universals are properties that can be predicated of any number of particulars. The traditional, Platonist account of universals is that they are real, abstract entities. They exist independently of particulars and are what general terms such as *yellow*, *umbrella*, and *cucumber* refer to. According to the Platonist, yellow things are yellow only because they are all related to the universal yellow. This view is unpopular with empiricists but has gained new currency with the advent of the universals theory of laws.

UNIVERSALS THEORY OF LAWS Because of the many difficulties afflicting the regularity theory, Fred Dretske, D. M. Armstrong, Michael Tooley, and others have proposed a necessitarian analysis of laws of nature. According to this analysis, laws are not universal generalizations but singular statements of fact about relations between universals—hence the name *universals theory* for this new treatment of laws.

VACUOUS LAWS Vacuous laws are not really laws at all: they are universal generalizations that are true solely because their antecedents are never satisfied. For example, since there are no mountains made of rhodium, it is true that all rhodium mountains are radioactive and also true that none of them are. The simple, Humean, regularity theory of laws insists that laws of nature are nothing more than true universal generalizations. Thus the regularity theorist has the problem of explaining why we do not accept vacuous laws as genuine. (See *noninstantial laws*, *regularity theory of laws*.)

VALIDITY (OF AN ARGUMENT) See *deduction*.

VALUE-NEUTRALITY THESIS The thesis that scientific decisions about theories should be governed exclusively by cognitive values. Cognitive (or epistemic) values are those values that are linked, directly or indirectly, to the aims of science as a knowledge-seeking enterprise, especially the goal of discovering interesting truths and rejecting error. Cognitive values include predictive power, explanatory scope, and (some would argue) simplicity.

VERIFIABILITY PRINCIPLE (OR CRITERION) OF MEANING As originally formulated by the logical positivists, the verifiability principle asserts that the meaning of any contingent statement is given by the observation statements needed to verify it conclusively. Observation statements are assumed to be directly verifiable by the experiences they purportedly describe. Unverifiable assertions are declared to be meaningless (or, at least, to lack any cognitive meaning). Criticisms by Karl Popper and others soon led to the abandonment of verifiability as a criterion of meaning in favor of weaker notions such as confirmability and testability. These, too, are controversial insofar as they are intended as explications of meaning.

VERISIMILITUDE The term used by Karl Popper and others to connote the objective truth content of a theory: the higher a theory's degree of verisimilitude, the closer it is to the truth. Attempts to define verisimilitude in terms of the ratio of the number of a theory's true consequences to the total number of the theory's consequences have run into enormous difficulties.

BIBLIOGRAPHY

CHAPTER 1 | SCIENCE AND PSEUDOSCIENCE

Bamford, Greg. "Popper and His Commentators on the Discovery of Neptune: A Close Shave for the Law of Gravitation?" *Studies in History and Philosophy of Science* 27 (1996): 207–32.

Brandon, Robert N. *Adaptation and Environment*. Princeton, N.J.: Princeton University Press, 1990.

Collins, Harry, and Trevor Pinch. *The Golem: What Everyone Should Know about Science*. Cambridge: Cambridge University Press, 1993.

Earman, John, and Clark Glymour. "Relativity and Eclipses: The British Eclipse Expeditions of 1919 and Their Predecessors." *Historical Studies in the Physical Sciences* 11 (1980): 49–85.

Gardner, Martin. *Fads and Fallacies in the Name of Science*. New York: Dover, 1957.

———. *Science: Good, Bad and Bogus*. Buffalo, N.Y.: Prometheus Books, 1981.

———. *The New Age: Notes of a Fringe Watcher*. Buffalo, N.Y.: Prometheus Books, 1991.

———. *On the Wild Side*. Buffalo, N.Y.: Prometheus Books, 1992.

Grim, Patrick, ed. *Philosophy of Science and the Occult*. Albany, N.Y.: SUNY Press, 1982.

Grosser, Morton. *The Discovery of Neptune*. Cambridge, Mass.: Harvard University Press, 1962.

Grove, Jack W. *In Defence of Science*. Toronto: University of Toronto Press, 1989.

Halstead, Beverly. "Popper: Good Philosophy, Bad Science?" *New Scientist* (17 July 1980): 215–17.

Kitcher, Philip. *Abusing Science: The Case against Creationism*. Cambridge, Mass.: MIT Press, 1982.

Lakatos, Imre. "Falsification and the Methodology of Scientific Research Programmes." In *Criticism and the Growth of Science*. Ed. Imre Lakatos and Alan Musgrave. Cambridge: Cambridge University Press, 1970.

Lakatos, Imre, and Alan Musgrave, eds. *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press, 1970.

Laudan, Larry. "More on Creationism." *Science, Technology, and Human Values* 8 (Winter 1983): 36–38.

———. "The Demise of the Demarcation Problem." In *Physics, Philosophy, and Psychoanalysis*. Ed. R. S. Cohen and L. Laudan. Dordrecht, Netherlands: D. Reidel, 1983.

Mayo, Deborah. "Novel Evidence and Severe Tests." *Philosophy of Science* 58 (1991): 523–52.

———. "Ducks, Rabbits, and Normal Science: Recasting the Kuhn's-eye View of Popper's Demarcation of Science." *British Journal for the Philosophy of Science* 47 (1996): 271–90.

Moore, John N., and Harold S. Slusher. *Biology: A Search for Order in Complexity*. Grand Rapids, Mich.: Zondervan Publishing House, 1970.

Morris, Henry M. *Scientific Creationism: The Public School Edition*. San Diego, Calif.: Creation-Life Publishers, 1974.

Pannekoek, Anton. *A History of Astronomy*, rpt. 1951; New York: Dover Publications, 1990.

Popper, Karl R. *The Logic of Scientific Discovery*. New York: Basic Books, 1959.

———. "Normal Science and Its Dangers." In *Criticism and the Growth of Science*. Ed. Imre Lakatos and Alan Musgrave. Cambridge: Cambridge University Press, 1970.

———. "Autobiography of Karl Popper." In *The Philosophy of Karl Popper*. Vol. 1. Ed. P. A. Schlipp. La Salle, Ill.: Open Court, 1974.

———. "Natural Selection and the Emergence of Mind." *Dialectica* 32 (1978): 339–55.

———. "Letter on Evolution." *New Scientist* (21 August 1980): 611.

Quinn, Philip L. "The Philosopher of Science as Expert Witness." In *Science and Reality*. Ed. J. T. Cushing, C. F. Delaney, and G. M. Gutting. Notre Dame, Ind.: University of Notre Dame Press, 1984.

———. "Creationism, Methodology, and Politics." In *But Is It Science?* Ed. Michael Ruse. Buffalo, N.Y.: Prometheus Books, 1988.

Ruse, Michael. *Darwinism Defended: A Guide to the Evolution Controversies*. Reading, Pa.: Addison-Wesley, 1982.

———. "The Academic as Expert Witness." *Science, Technology, and Human Values* 11 (Spring 1986): 68–73.

Ruse, Michael, ed. *But Is It Science?* Buffalo, N.Y.: Prometheus Books, 1988.

Schick, Theodore, Jr., and Lewis Vaughn. *How to Think about Weird Things: Critical Thinking for a New Age*. Mountain View, Calif.: Mayfield, 1995.

Sober, Elliott. *The Nature of Selection*. Chicago: University of Chicago Press, 1984.

———. *Philosophy of Biology*. Boulder, Colo.: Westview Press, 1993.

Thagard, Paul. "Pseudoscience." Chap. 9, *Computational Philosophy of Science*. Cambridge, Mass.: MIT Press, 1988.

Thompson, R. Paul. "Is Sociobiology a Pseudoscience?" In *PSA 1980*. Vol. 1. Ed. P. D. Asquith and R. N. Giere. East Lansing, Mich.: Philosophy of Science Association, 1980.

CHAPTER 2 | RATIONALITY, OBJECTIVITY, AND VALUES IN SCIENCE

Audi, Robert. "Scientific Objectivity and the Evaluation of Hypotheses." In *The Philosophy of Logical Mechanism*. Ed. M. H. Salmon. Dordrecht, Netherlands: Kluwer Academic Publishers, 1990.

Brown, Harold I. *Perception, Theory, and Commitment: The New Philosophy of Science*. Chicago: University of Chicago Press, 1977.

———. "Incommensurability." *Inquiry* 26 (1983): 3–29.

———. "Response to Siegel." *Synthese* 56 (1983): 91–105.

Cedarbaum, Daniel Goldman. "Paradigms." *Studies in History and Philosophy of Science* 14 (1983): 173–213.

Curd, Martin. "Kuhn, Scientific Revolutions, and the Copernican Revolution." *Nature and System* 6 (1984): 1–14.

Doppelt, Gerald. "Kuhn's Epistemological Relativism: An Interpretation and Defense." *Inquiry* 21 (1978): 33–86.

———. "A Reply to Siegel on Kuhnian Relativism." *Inquiry* 23 (1980): 117–23.

———. "Relativism and the Reticulational Model of Scientific Rationality." *Synthese* 69 (1986): 225–52.

Glymour, Clark. *Theory and Evidence*. Princeton, N.J.: Princeton University Press, 1980.

Gutting, Gary, ed. *Paradigms and Revolutions*. Notre Dame, Ind.: University of Notre Dame Press, 1980.

Harding, Sandra. *Whose Science? Whose Knowledge? Thinking from Women's Lives*. Ithaca, N.Y.: Cornell University Press, 1991.

Harding, Sandra, and Jean F. O'Barr, eds. *Sex and Scientific Inquiry*. Chicago: University of Chicago Press, 1987.

Horwich, Paul, ed. *World Changes: Thomas Kuhn and the Nature of Science*. Cambridge, Mass.: MIT Press, 1993.

Hoyningen-Huene, Paul. *Reconstructing Scientific Revolutions: Thomas Kuhn's Philosophy of Science*. Chicago: University of Chicago Press, 1993.

Kordig, Carl R. "The Theory-Ladenness of Observation." *Review of Metaphysics* 24 (1971): 448–84.

———. *The Justification of Scientific Change*. Dordrecht, Netherlands: D. Reidel, 1971.

Kourany, Janet A. "The Nonhistorical Basis of Kuhn's Theory of Science." *Nature and System* 1 (1979): 46–59.

Kuhn, Thomas S. *The Essential Tension*. Chicago: University of Chicago Press, 1977.

———. "Rationality and Theory Choice." *Journal of Philosophy* 80 (1983): 563–70.

———. *The Structure of Scientific Revolutions*, 3d ed. Chicago: University of Chicago Press, 1996.

Lakatos, Imre, and Alan Musgrave, eds. *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press, 1970.

Laslett, Barbara, et al., eds. *Gender and Scientific Authority*. Chicago: University of Chicago Press, 1996.

Laudan, Larry. *Science and Values*. Berkeley, Calif.: University of California Press, 1984.

———. "Relativism, Naturalism, and Reticulation." *Synthese* 71 (1987): 221–34.

———. *Science and Relativism: Some Key Controversies in the Philosophy of Science*. Chicago: University of Chicago Press, 1990.

Longino, Helen E. "Beyond 'Bad Science': Skeptical Reflections on the Value-Freedom of Scientific Inquiry." *Science, Technology, and Human Values* 8 (1983): 7–17.

———. "Can There Be a Feminist Science?" *Hypatia* 2 (1987): 51–64.

———. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton, N.J.: Princeton University Press, 1990.

———. "In Search of Feminist Epistemology." *The Monist* 77 (1994): 427–85.

Martin, Jane R. "Ideological Critiques and the Philosophy of Science." *Philosophy of Science* 56 (1989): 1–22.

McMullin, Eran. "Values in Science." In *PSA 1982*. Vol. 2. Ed. P. D. Asquith and T. Nickles. East Lansing, Mich.: Philosophy of Science Association, 1982.

Musgrave, Alan. "Kuhn's Second Thoughts." Review of *The Structure of Scientific Revolutions* 2d ed., by Thomas S. Kuhn. *British Journal for the Philosophy of Science* 22 (1971): 287–97. Reprinted in Gutting, Gary, ed. *Paradigms and Revolutions*. Notre Dame, Ind.: University of Notre Dame Press, 1980.

Pinnick, Cassandra. "Feminist Epistemology: Implications for Philosophy of Science." *Philosophy of Science* 61 (1994): 646–57.

Richardson, Robert C. "Biology and Ideology: The Interpenetration of Science and Values." *Philosophy of Science* 51 (1984): 396–420.

Salmon, Wesley C. "Rationality and Objectivity in Science, or Tom Kuhn Meets Tom Bayes." In *Scientific Theories*. Ed. C. W. Savage. Vol. 14, *Minnesota Studies in the Philosophy of Science*. Minneapolis: University of Minnesota Press, 1990.

Scheffler, Israel. *Science and Subjectivity*. Indianapolis, Ind.: Bobbs-Merrill, 1967.

Shapere, Dudley. Review of *The Structure of Scientific Revolutions* by Thomas S. Kuhn. *Philosophical Review* 73 (1964): 383–94. Reprinted in Gutting, Gary, ed. *Paradigms and Revolutions*. Notre Dame, Ind.: University of Notre Dame Press, 1980.

———. "The Paradigm Concept." *Science* 172 (1971): 706–9.

Shimony, Abner. "Comments on Two Epistemological Theses of Thomas Kuhn." In *Essays in Memory of Imre Lakatos*. Ed. R. S. Cohen et al. Dordrecht, Netherlands: D. Reidel, 1976. Reprinted in Abner Shimony, *Search for a Naturalistic World View*, Vol. 1, *Scientific Method and Epistemology*. Cambridge: Cambridge University Press, 1993.

Siegel, Harvey. "Epistemological Relativism in its Latest Form." *Inquiry* 23 (1980): 107-17.

———. "Brown on Epistemology and the New Philosophy of Science." *Synthese* 56 (1983): 61-89.

CHAPTER 3 | THE DUHEM-QUINE THESIS AND UNDERDETERMINATION

Barker, Stephen F. "Logical Positivism and the Philosophy of Mathematics." In *The Legacy of Logical Positivism*. Ed. P. Achinstein and S. F. Barker. Baltimore, Md.: The Johns Hopkins University Press, 1969.

Barker, Stephen F., and Peter Achinstein. "On the New Riddle of Induction." *Philosophical Review* 69 (1960): 511-22.

Bloor, David. "The Strengths of the Strong Programme." *Philosophy of the Social Sciences* 11 (1981): 199-213.

———. "Durkheim and Mauss Revisited: Classification and the Sociology of Knowledge." *Studies in History and Philosophy of Science* 13 (1982): 267-97.

———. "A Reply to Gerd Buchdahl." *Studies in History and Philosophy of Science* 13 (1982): 305-11.

———. *Knowledge and Social Imagery*, 2d ed. Chicago: University of Chicago Press, 1991.

BonJour, Laurence. "A Rationalist Manifesto." In *Return of the A Priori*. Ed. P. Hanson and B. Hunter. Suppl. vol. 18, *Canadian Journal of Philosophy*. Calgary, Alberta: University of Calgary Press, 1992.

Brown, James R., ed. *Scientific Rationality: The Sociological Turn*. Dordrecht, Netherlands: D. Reidel, 1984.

Buchdahl, Gerd. "Editorial Response to David Bloor." *Studies in History and Philosophy of Science* 13 (1982): 299-304.

Collingwood, R. G. *The Idea of History*. Oxford: Clarendon Press, 1946.

Duhem, Pierre. *The Aim and Structure of Physical Theory*. Trans. Philip P. Wiener. Princeton, N.J.: Princeton University Press, 1954.

Fine, Arthur. "Science Made Up: Constructivist Sociology of Scientific Knowledge." In *The Disunity of Science*. Ed. P. Galison and D. J. Stump. Stanford, Calif.: Stanford University Press, 1996.

Gibbins, Peter. *Particles and Paradoxes: The Limits of Quantum Logic*. Cambridge: Cambridge University Press, 1987.

Glymour, Clark. *Theory and Evidence*. Princeton, N.J.: Princeton University Press, 1980.

Goodman, Nelson. "Positionality and Pictures." *Philosophical Review* 69 (1960): 523-25.

———. *Fact, Fiction, and Forecast*, 4th ed. Cambridge, Mass.: Harvard University Press, 1983.

Grice, H. P., and P. F. Strawson. "In Defense of a Dogma." *Philosophical Review* 65 (1956): 141-58.

Haack, Susan. *Philosophy of Logics*. Cambridge: Cambridge University Press, 1978.

———. *Deviant Logic, Fuzzy Logic*. Chicago: University of Chicago Press, 1996.

Harding, Sandra, ed., *Can Theories Be Refuted? Essays on the Duhem-Quine Thesis*. Dordrecht, Netherlands: D. Reidel, 1976.

Hausman, Alan. "Non-Euclidean Geometry and Relative Consistency Proofs." In *Motion and Time, Space and Matter*. Ed. P. K. Machamer and R. G. Turnbull. Columbus: Ohio State University Press, 1976.

Hempel, Carl G. "Geometry and Empirical Science." *American Mathematical Monthly* 52 (1945): 7–17.

———. "On the Nature of Mathematical Truth." *American Mathematical Monthly* 52 (1945): 543–56.

Hesse, Mary B. "The Strong Thesis of Sociology of Science." In *Revolutions and Reconstructions in the Philosophy of Science*. Bloomington: Indiana University Press, 1980.

Hoefer, Carl, and Alexander Rosenberg. "Empirical Equivalence, Underdetermination, and Systems of the World." *Philosophy of Science* 61 (1994): 592–607.

Horwich, Paul. *Probability and Evidence*. Cambridge: Cambridge University Press, 1982.

Howson, Colin, and Peter Urbach. *Scientific Reasoning: The Bayesian Approach*, 2d ed. La Salle, Ill.: Open Court, 1993.

Kukla, André. "Laudan, Leplin, Empirical Equivalence, and Underdetermination." *Analysis* 53 (1993): 1–7.

Lakatos, Imre. "History of Science and Its Rational Reconstructions." In *Method and Appraisal in the Physical Sciences*. Ed. C. Howson. Cambridge: Cambridge University Press, 1976.

Laudan, Larry. "Grünbaum on 'The Duhemian Argument.'" *Philosophy of Science* 32 (1965): 295–99.

———. "The Pseudo-Science of Science?" *Philosophy of the Social Sciences* 11 (1981): 173–98.

———. "More on Bloor." *Philosophy of the Social Sciences* 12 (1982): 71–74.

Laudan, Larry, and Jarrett Leplin. "Empirical Equivalence and Underdetermination." *Journal of Philosophy* 88 (1991): 449–72.

Laymon, Ronald. "Newton's Demonstration of Universal Gravitation and Philosophical Theories of Confirmation." In *Testing Scientific Theories*. Ed. J. Earman. Vol. 10, *Minnesota Studies in the Philosophy of Science*. Minneapolis: University of Minnesota Press, 1983.

Leplin, Jarrett, and Larry Laudan. "Determination Undeterred: Reply to Kukla." *Analysis* 53 (1993): 8–16.

McCarthy, Thomas. "Scientific Rationality and the 'Strong Program' in the Sociology of Knowledge." In *Construction and Constraint*. Ed. E. McMullin. Notre Dame, Ind.: University of Notre Dame Press, 1988.

- Musgrave, Alan. "Logicism Revisited." *British Journal for the Philosophy of Science* 28 (1977): 99–127.
- Nagel, Ernest. *The Structure of Science*. London: Routledge and Kegan Paul, 1961.
- Putnam, Hilary. "Is Logical Empirical?" in *Boston Studies in the Philosophy of Science*. Vol. 5. Ed. R. S. Cohen and M. W. Wartofsky. Dordrecht, Netherlands: D. Reidel, 1968. Revised version reprinted as "The Logic of Quantum Mechanics." In *Mathematics, Matter, and Method, Philosophical Papers*. Vol. 1. Cambridge: Cambridge University Press, 1975.
- . "'Two Dogmas' Revisited." In *Contemporary Aspects of Philosophy*. Ed. G. Ryle. Boston, Mass.: Oriol Press, 1976. Reprinted in *Realism and Reason, Philosophical Papers*. Vol. 3. Cambridge: Cambridge University Press, 1983.
- . "There Is At Least One A Priori Truth." *Erkenntnis* 13 (1978): 153–70. Reprinted in *Realism and Reason, Philosophical Papers*. Vol. 3. Cambridge: Cambridge University Press, 1983.
- Quine, Willard V. *Philosophy of Logic*. Englewood Cliffs, N.J.: Prentice-Hall, 1970.
- . "On Empirically Equivalent Systems of the World." *Erkenntnis* 9 (1975): 313–28.
- . "Two Dogmas in Retrospect." *Canadian Journal of Philosophy* 21 (1991): 265–74.
- Quinn, Philip. "What Duhem Really Meant." In *Methodological and Historical Essays in the Natural and Social Sciences*. Ed. R. S. Cohen and M. W. Wartofsky. Dordrecht, Netherlands: D. Reidel, 1973.
- . "Rejoinder to Tuana." *Philosophy of Science* 45 (1978): 463–65.
- Rosenkrantz, Roger D. "Does the Philosophy of Induction Rest on a Mistake?" *Journal of Philosophy* 79 (1982): 78–97.
- Slezak, Peter. "The Social Construction of Social Constructionism." *Inquiry* 37 (1994): 139–57.
- . "A Second Look at David Bloor's *Knowledge and Social Imagery*." *Philosophy of the Social Sciences* 24 (1994): 336–61.
- Stalker, Douglas, ed. *Grue! The New Riddle of Induction*. Chicago: Open Court, 1994.
- Tuana, Nancy. "Quinn on Duhem: An Emendation." *Philosophy of Science* 45 (1978): 456–62.

CHAPTER 4 | INDUCTION, PREDICTION, AND EVIDENCE

- Achinstein, Peter. "The Method of Hypothesis: What Is It Supposed to Do, and Can It Do It?" In *Observation, Experiment, and Hypothesis in Modern Physical Science*. Ed. Peter Achinstein and Owen Hannaway. Cambridge, Mass.: MIT Press, 1985.
- . *Particles and Waves*. New York: Oxford University Press, 1991.
- . "Are Empirical Evidence Claims A Priori?" *British Journal for the Philosophy of Science* 46 (1995): 447–73.

Achinstein, Peter, ed. *The Concept of Evidence*. Oxford: Oxford University Press, 1983.

Achinstein, Peter, and Owen Hannaway, eds. *Observation, Experiment, and Hypothesis in Modern Physical Science*. Cambridge, Mass.: MIT Press, 1985.

Barker, Stephen F. *Induction and Hypothesis: A Study of the Logic of Confirmation*. Ithaca, N.Y.: Cornell University Press, 1957.

Brush, Stephen G. "Prediction and Theory Evaluation: The Case of Light Bending." *Science* 246 (1989): 1124–29.

———. "Dynamics of Theory Change: The Role of Predictions." In *PSA 1994*. Vol. 2. Ed. D. Hull, M. Forbes, and R. M. Burian. East Lansing, Mich.: Philosophy of Science Association, 1994.

Chalmers, Alan F. "On Learning from Our Mistakes." *British Journal for the Philosophy of Science* 24 (1973): 164–73.

Collins, Robin. "Against the Epistemic Value of Prediction over Accommodation." *Noûs* 28 (1994): 210–24.

Curd, Martin. "The Logic of Discovery: An Analysis of Three Approaches." In *Scientific Discovery, Logic, and Rationality*. Ed. Thomas Nickles. Dordrecht, Netherlands: D. Reidel, 1980.

Erwin, Edward, and Harvey Siegel. "Is Confirmation Differential?" *British Journal for the Philosophy of Science* 40 (1989): 105–19.

Fumerton, Richard A. "Induction and Reasoning to the Best Explanation." *Philosophy of Science* 47 (1980): 589–600.

Gardner, Michael R. "Predicting Novel Facts," *British Journal for the Philosophy of Science* 33 (1982): 1–15.

Gingerich, Owen. "Neptune, Velikovsky, and the Name of the Game." *Scientific American* 275 (September 1996): 181–83.

Glymour, Clark. "Hypothetico-Deductivism Is Hopeless." *Philosophy of Science* 47 (1980): 322–25.

———. *Theory and Evidence*. Princeton, N.J.: Princeton University Press, 1980.

Gooding, David, Trevor Pinch, and Simon Schaffer, eds. *The Uses of Experiment*. Cambridge: Cambridge University Press, 1989.

Grünbaum, Adolf. "Is the Method of Bold Conjectures and Attempted Refutations Justifiably the Method of Science?" *British Journal for the Philosophy of Science* 27 (1976): 105–36.

Haack, Susan. "The Justification of Induction." *Mind* 85 (1976): 112–19.

Harman, Gilbert. "The Inference to the Best Explanation." *Philosophical Review* 74 (1965): 88–95.

Hempel, Carl G. "Studies in the Logic of Confirmation," *Mind* 54 (1945): 1–26, 97–121. Reprinted in *Aspects of Scientific Explanation*. New York: Macmillan, 1965.

———. *Philosophy of Natural Science*. Englewood Cliffs, N.J.: Prentice-Hall, 1966.

Horwich, Paul. *Probability and Evidence*. Cambridge: Cambridge University Press, 1982.

Jeffrey, Richard C. "Probability and the Art of Judgment," in *Observation, Experiment, and Hypothesis in Modern Physical Science*. Ed. Peter Achinstein and Owen Hannaway. Cambridge, Mass.: MIT Press, 1985.

Laudan, Larry. "Why Was the Logic of Discovery Abandoned?" In *Scientific Discovery, Logic, and Rationality*. Ed. Thomas Nickles. Dordrecht, Netherlands: D. Reidel, 1980.

Leplin, Jarrett. "The Bearing of Discovery on Justification." *Canadian Journal of Philosophy* 17 (1987): 805–14.

Lipton, Peter. "Prediction and Prejudice." *International Studies in the Philosophy of Science* 4 (1990): 51–65. Reprinted in *Inference to the Best Explanation*. New York: Routledge, 1991.

———. *Inference to the Best Explanation*. New York: Routledge, 1991.

Mayo, Deborah G. "Novel Evidence and Severe Tests." *Philosophy of Science* 58 (1991): 523–52.

Mellor, D. H. "The Warrant of Induction." In *Matters of Metaphysics*. Cambridge: University of Cambridge Press, 1991.

Miller, David. "Can Science Do Without Induction?" In *Applications of Inductive Logic*. Ed. L. J. Cohen and M. B. Hesse. Oxford: Clarendon Press, 1980.

Musgrave, Alan. "Logical versus Historical Theories of Confirmation." *British Journal for the Philosophy of Science* 25 (1974): 1–23.

Newton-Smith, W. H. *The Rationality of Science*. Boston, Mass.: Routledge and Kegan Paul, 1981.

Nagel, Ernest. "Carnap's Theory of Induction." In *The Philosophy of Rudolf Carnap*. Ed. P. A. Schlipp. LaSalle, Ill.: Open Court, 1963.

Nickles, Thomas. "Justification and Experiment." In *The Uses of Experiment*. Ed. David Gooding, Trevor Pinch, and Simon Schaffer. Cambridge: Cambridge University Press, 1989.

Nickles, Thomas, ed., *Scientific Discovery, Logic, and Rationality*. Dordrecht, Netherlands: D. Reidel, 1980.

Popper, Karl R. *The Logic of Scientific Discovery*. New York: Basic Books, 1959.

———. *Conjectures and Refutations*. New York: Harper and Row, 1963.

Rappaport, Steven. "Inference to the Best Explanation: Is It Really Different from Mill's Methods?" *Philosophy of Science* 63 (1996): 65–80.

Reichenbach, Hans. *Experience and Prediction*. Chicago: University of Chicago Press, 1938.

Russow, Lilly-Marlene, and Martin Curd. *Principles of Reasoning*. New York: St. Martin's Press, 1989.

Salmon, Wesley C. *The Foundations of Scientific Inference*. Pittsburgh, Pa.: University of Pittsburgh Press, 1967.

———. "Unfinished Business: The Problem of Induction." *Philosophical Studies* 33 (1978): 1–19.

Skyrms, Brian. *Choice and Chance: An Introduction to Inductive Logic*, 3d ed. Belmont, Calif.: Wadsworth, 1986.

Swinburne, Richard G., ed. *The Justification of Induction*. Oxford: Oxford University Press, 1974.

Van Cleve, James. "Reliability, Justification, and the Problem of Induction." In *Midwest Studies in Philosophy*. Vol. 9. Ed. P. French, T. Uehling, Jr., and H. Wettstein. Minneapolis: University of Minnesota Press, 1984.

Watkins, John. *Science and Scepticism*. Princeton, N.J.: Princeton University Press, 1984.

Worrall, John. "Fresnel, Poisson, and the White Spot: The Role of Successful Predictions in the Acceptance of Scientific Theories." In *The Uses of Experiment*. Ed. David Gooding, Trevor Pinch, and Simon Schaffer. Cambridge: Cambridge University Press, 1989.

CHAPTER 5 | CONFIRMATION AND RELEVANCE: BAYESIAN APPROACHES

Berger, James O., and Donald A. Berry. "Statistical Analysis and the Illusion of Objectivity." *American Scientist* 76 (1988): 159–65.

Brown, Harold I. "Reason, Judgement and Bayes's Law." *Philosophy of Science* 61 (1994): 351–69.

Dorling, Jon. "Bayesian Personalism, the Methodology of Scientific Research Programmes, and Duhem's Problem." *Studies in History and Philosophy of Science* 10 (1979): 177–87.

Earman, John. *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge, Mass.: MIT Press, 1992.

Earman, John, ed. *Testing Scientific Theories*. Vol. 10, *Minnesota Studies in the Philosophy of Science*. Minneapolis: University of Minnesota Press, 1983.

Forster, Malcolm. "Bayes and Bust: Simplicity as a Problem for a Probabilist's Approach to Confirmation." *British Journal for the Philosophy of Science* 46 (1995): 399–424.

Garber, Daniel. "Old Evidence and Logical Omniscience in Bayesian Confirmation Theory." In *Testing Scientific Theories*. Ed. John Earman. Vol. 10, *Minnesota Studies in the Philosophy of Science*. Minneapolis: University of Minnesota Press, 1983.

Good, I. J. "The White Shoe Is a Red Herring." *British Journal for the Philosophy of Science* 17 (1967): 322.

Goodman, Nelson. *Fact, Fiction, and Forecast*, 4th ed. Cambridge, Mass.: Harvard University Press, 1983.

Hempel, Carl G. "Studies in the Logic of Confirmation." *Mind* 54 (1945): 1–26, 97–121. Reprinted in *Aspects of Scientific Explanation*. New York: Macmillan, 1965.

Hesse, Mary B. *The Structure of Scientific Inference*. Berkeley, Calif.: University of California Press, 1974.

Horwich, Paul. "An Appraisal of Glymour's Confirmation Theory." *Journal of Philosophy* 75 (1978): 98–113.

———. *Probability and Evidence*. Cambridge: University of Cambridge Press, 1982.

- Howson, Colin. "Theories of Probability." *British Journal for the Philosophy of Science* 46 (1995): 1–32.
- Howson, Colin, and Peter Urbach. *Scientific Reasoning: The Bayesian Approach*, 2d ed. La Salle, Ill.: Open Court, 1993.
- Jeffrey, Richard C. *The Logic of Decision*, 2d ed. Chicago: University of Chicago Press, 1983.
- . "Bayesianism with a Human Face." In *Testing Scientific Theories*. Ed. John Earman. Vol. 10, *Minnesota Studies in the Philosophy of Science*. Minneapolis: University of Minnesota Press, 1983.
- Kyburg, Henry E., Jr. *Probability and Inductive Logic*. London: Collier-Macmillan, 1970.
- . "Subjective Probability: Criticisms, Reflections, and Problems." *Journal of Philosophical Logic* 7 (1978): 157–80.
- Mackie, J. L. "The Paradox of Confirmation." *British Journal for the Philosophy of Science* 13 (1963): 265–77.
- . "The Relevance Criterion of Confirmation." *British Journal for the Philosophy of Science* 20 (1969): 27–40.
- Mayo, Deborah G. *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press, 1996.
- Nagel, Ernest. "Carnap's Theory of Induction." In *The Philosophy of Rudolf Carnap*. Ed. P. A. Schlipp. LaSalle, Ill.: Open Court, 1963.
- Nelson, David E. "Confirmation, Explanation, and Logical Strength." *British Journal for the Philosophy of Science* 47 (1996): 399–413.
- Niiniluoto, Ilkka. "Novel Facts and Bayesianism." *British Journal for the Philosophy of Science* 34 (1983): 375–79.
- Plantinga, Alvin. "The Probabilistic Argument from Evil." *Philosophical Studies* 35 (1979): 1–53.
- . "Bayesian Coherentism and Warrant" and "Bayesian Coherentism and Rationality." Chaps. 6 and 7, *Warrant: The Current Debate*. Oxford: University of Oxford Press, 1993.
- . "Epistemic Probability: Some Current Views" and "Epistemic Conditional Probability: The Sober Truth." Chaps. 8 and 9, *Warrant and Proper Function*. Oxford: University of Oxford Press, 1993.
- Rosenkrantz, Roger D. *Inference, Method and Decision*. Dordrecht, Netherlands: D. Reidel, 1977.
- . "Does the Philosophy of Induction Rest on a Mistake?" *Journal of Philosophy* 79 (1982): 78–97.
- . "Why Glymour Is a Bayesian." In *Testing Scientific Theories*. Ed. John Earman. Vol. 10, *Minnesota Studies in the Philosophy of Science*. Minneapolis: University of Minnesota Press, 1983.
- Salmon, Wesley C. *The Foundations of Scientific Inference*. Pittsburgh, Pa.: University of Pittsburgh Press, 1967.
- . "Confirmation." *Scientific American* 228 (May 1973): 75–83.

———. "Confirmation and Relevance." In *Induction, Probability, and Confirmation*. Vol. 6, *Minnesota Studies in the Philosophy of Science*. Ed. G. Maxwell and R. M. Anderson, Jr. Minneapolis: University of Minnesota Press, 1975.

———. "The Appraisal of Theories: Kuhn Meets Bayes." In *PSA 1990*. Vol. 2. Ed. A. Fine, M. Forbes, and L. Wessels. East Lansing, Mich.: Philosophy of Science Association, 1991.

Steel, Daniel. "Bayesianism and the Value of Diverse Evidence." *Philosophy of Science* 63 (1996): 666–74.

Swinburne, Richard G. "The Paradoxes of Confirmation—A Survey." *American Philosophical Quarterly* 8 (1971): 318–30.

———. *An Introduction to Confirmation Theory*. (London: Methuen and Co., 1973).

Teller, Paul. "Conditionalization and Observation." *Synthese* 26 (1973): 218–58.

Vineberg, Susan. "Eliminative Induction and Bayesian Confirmation Theory." *Canadian Journal of Philosophy* 26 (1996): 257–66.

Wavne, Andrew. "Bayesianism and Diverse Evidence." *Philosophy of Science* 62 (1995): 111–21.

CHAPTER 6 | MODELS OF EXPLANATION

Achinstein, Peter. "Can There Be a Model of Explanation?" *Theory and Decision* 13 (1981): 201–27.

Brody, Baruch. "Towards an Aristotelian Theory of Scientific Explanation." *Philosophy of Science* 39 (1972): 20–31.

Coffa, J. Alberto. "Hempel's Ambiguity." *Synthese* 28 (1974): 141–63.

Donagan, Alan. "The Popper-Hempel Theory Reconsidered." In *Philosophical Analysis and History*. Ed. W. H. Dray. New York: Harper and Row, 1966.

Dray, William H. *Laws and Explanation in History*. Oxford: Oxford University Press, 1957.

Gardiner, Patrick L., ed. *Theories of History*. New York: Free Press, 1959.

Hempel, Carl G. "The Function of General Laws in History." *Journal of Philosophy* 39 (1942): 35–42. Reprinted in *Aspects of Scientific Explanation*. New York: Free Press, 1965.

———. "Deductive-Nomological vs. Statistical Explanation." In *Scientific Explanation, Space, and Time*. Vol. 3, *Minnesota Studies in the Philosophy of Science*. Ed. H. Feigl and G. Maxwell. Minneapolis: University of Minnesota Press, 1962.

———. "Aspects of Scientific Explanation." In *Aspects of Scientific Explanation*. New York: Free Press, 1965.

———. "Maximal Specificity and Lawlikeness in Probabilistic Explanation." *Philosophy of Science* 35 (1968): 116–33.

Hempel, Carl G., and Paul Oppenheim. "Studies in the Logic of Explanation." *Philosophy of Science* 15 (1948): 567–79. Reprinted in *Aspects of Scientific Explanation*. New York: Free Press, 1965.

Humphreys, Paul. "Why Propensities Cannot Be Probabilities." *Philosophical Review* 94 (1985): 557–70.

Kitcher, Philip, and Wesley C. Salmon, eds. *Scientific Explanation*. Vol. 13, *Minnesota Studies in the Philosophy of Science*. Minneapolis: University of Minnesota Press, 1989.

McCarthy, Timothy. "Discussion on an Aristotelian Model of Scientific Explanation." *Philosophy of Science* 44 (1977): 159–66.

Meixner, J. "Homogeneity and Explanatory Depth." *Philosophy of Science* 46 (1979): 366–81.

Papineau, David. *For Science in the Social Sciences*. New York: St. Martin's Press, 1978.

Pitt, Joseph C., ed. *Theories of Explanation*. Oxford: Oxford University Press, 1988.

Ruben, David-Hillel. *Explaining Explanation*. New York: Routledge, 1990.

Ruben, David-Hillel, ed. *Explanation*. Oxford: Oxford University Press, 1993.

Salmon, Merrilee H. "Explanation in the Social Sciences." In *Scientific Explanation*. Ed. Philip Kitcher and Wesley C. Salmon. Vol. 13, *Minnesota Studies in the Philosophy of Science*. Minneapolis: University of Minnesota Press, 1989.

Salmon, Wesley C. *Statistical Explanation and Statistical Relevance*. Pittsburgh, Pa. University of Pittsburgh Press, 1971.

———. *Scientific Explanation and the Causal Structure of the World*. Princeton, N.J.: Princeton University Press, 1984.

———. "Four Decades of Scientific Explanation." In *Scientific Explanation*. Ed. Philip Kitcher and Wesley C. Salmon. Vol. 13, *Minnesota Studies in the Philosophy of Science*. Minneapolis: University of Minnesota Press, 1989. Also published as *Four Decades of Scientific Explanation*. Minneapolis: University of Minnesota Press, 1989.

Scriven, Michael. "Explanation and Prediction in Evolutionary Theory." *Science* 130 (1959): 477–82.

———. "Tautisms as the Grounds for Historical Explanations." In *Theories of History*. Ed. Patrick L. Gardiner. New York: Free Press, 1959.

CHAPTER 7 | LAWS OF NATURE

Armstrong, D. M. "Laws of Nature as Relations between Universals, and as Universals." *Philosophical Topics* 13 (1982): 7–24.

———. *What Is a Law of Nature?* Cambridge: Cambridge University Press, 1983.

Beauchamp, Tom L., ed. *Philosophical Problems of Causation*. Belmont, Calif.: Dickenson, 1974.

Braithwaite, Richard B. *Scientific Explanation*. Cambridge: Cambridge University Press, 1953.

Carroll, John W. *Laws of Nature*. Cambridge: Cambridge University Press, 1994.

Cartwright, Nancy. *How the Laws of Physics Lie*. Oxford: Clarendon Press, 1983.

———. *Nature's Capacities and Their Measurement*. Oxford: Clarendon Press, 1989.

Chalmers, Alan. "So the Laws of Physics Needn't Lie." *Australasian Journal of Philosophy* 71 (1993): 196–205.

———. "Cartwright on Fundamental Laws: A Response to Clarke." *Australasian Journal of Philosophy* 74 (1996): 150–52.

Chisholm, Roderick M. "Law Statements and Counterfactual Inference." *Analysis* 15 (1955): 97–105.

Clarke, Steve. "The Lies Remain the Same: A Reply to Chalmers." *Australasian Journal of Philosophy* 73 (1995): 152–55.

Creary, Lewis. "Causal Explanation and the Reality of Natural Component Forces." *Pacific Philosophical Quarterly* 62 (1981): 148–57.

Donnellan, Keith S. "Kripke and Putnam on Natural Kind Terms." In *Knowledge and Mind*. Ed. C. Ginet and S. Shoemaker. Oxford: Oxford University Press, 1983.

Dretske, Fred I. "Reply to Niiniluoto." *Philosophy of Science* 45 (1978): 440–44.

Feynman, Richard P. *The Character of Physical Law*. Cambridge, Mass.: MIT Press, 1965.

Giere, Ronald N. "The Skeptical Perspective: Science without Laws of Nature." In *Laws of Nature: Essays on the Philosophical, Scientific, and Historical Dimensions*. Ed. Friedel Weinert. New York: Walter de Gruyter, 1995.

Hempel, Carl G. "Provisos: A Problem Concerning the Inferential Function of Scientific Theories." In *The Limitations of Deductivism*. Ed. A. Grünbaum and W. C. Salmon. Berkeley, Calif.: University of California Press, 1988.

Kline, A. David, and Carl A. Matheson. "How the Laws of Physics Don't Even Fib." In *PSA 1986*. Vol. 1. Ed. A. Fine and P. Machamer. East Lansing, Mich.: Philosophy of Science Association, 1986.

Kneale, William C. "Natural Laws and Contrary-to-Fact Conditionals." *Analysis* 10 (1950): 121–25. Reprinted in *Philosophical Problems of Causation*. Ed. Tom L. Beauchamp. Belmont, Calif.: Dickenson, 1974.

Kripke, Saul A. *Naming and Necessity*. Cambridge, Mass.: Harvard University Press, 1972.

Laymon, Ronald. "Cartwright and the Lying Laws of Physics." *Journal of Philosophy* 86 (1989): 353–72.

Levin, Michael. "Rigid Designators: Two Applications." *Philosophy of Science* 54 (1987): 283–94.

Lewis, David. "New Work for the Theory of Universals." *Australasian Journal of Philosophy* 61 (1983): 343–77.

Mackie, J. L. "Counterfactuals and Causal Laws." In *Analytical Philosophy: First Series*. Ed. R. J. Butler. Oxford: Basil Blackwell, 1966. Reprinted in *Philosophical Problems of Causation*. Ed. Tom L. Beauchamp. Belmont, Calif.: Dickenson, 1974.

Mellor, D. H. "Natural Kinds." *British Journal for the Philosophy of Science* 28 (1977): 299–312.

Molnar, George. "Kneale's Argument Revisited." *Philosophical Review* 78 (1969): 79–89.

Musgrave, Alan. "Wittgensteinian Instrumentalism." *Theoria* 47 (1981): 65–105.

- Nagel, Ernest. "The Logical Character of Scientific Laws." Chap. 4, *The Structure of Science*. New York: Harcourt, Brace and World, 1961.
- Niiniluoto, Ilkka. "Dretske on Laws of Nature." *Philosophy of Science* 45 (1978): 431–39.
- Schwartz, Stephen P., ed. *Naming, Necessity, and Natural Kinds*. Ithaca, N.Y.: Cornell University Press, 1977.
- Suchting, W. A. "Regularity and Law." In *Boston Studies in the Philosophy of Science*. Vol. 14. Ed. R. S. Cohen and M. W. Wartofsky. Dordrecht, Netherlands: D. Reidel, 1974.
- Swartz, Norman. *The Concept of Physical Law*. Cambridge: Cambridge University Press, 1985.
- . "A Neo-Humean Perspective: Laws as Regularities." In *Laws of Nature: Essays on the Philosophical, Scientific, and Historical Dimensions*. Ed. Friedel Weinert. New York: Walter de Gruyter, 1995.
- Swoyer, Christopher. "The Nature of Natural Laws." *Australasian Journal of Philosophy* 60 (1982): 203–23.
- Tooley, Michael. "The Nature of Laws." *Canadian Journal of Philosophy* 7 (1977): 667–98.
- Tweedale, Martin. "Universals and Laws of Nature." *Philosophical Topics* 13 (1982): 25–44.
- Urbach, Peter. "What Is a Law of Nature? A Humean Answer." *British Journal for the Philosophy of Science* 39 (1988): 193–210.
- Van Fraassen, Bas C. *Laws and Symmetry*. Oxford: Clarendon Press, 1989.
- Weinert, Friedel, ed. *Laws of Nature: Essays on the Philosophical, Scientific, and Historical Dimensions*. New York: Walter de Gruyter, 1995.
- Woodward, James. "Realism about Laws." *Erkenntnis* 36 (1992): 181–218.

CHAPTER 8 | INTERTHEORETIC REDUCTION

- Feyerabend, Paul. "Explanation, Reduction, and Empiricism." In *Scientific Explanation, Space, and Time*. Ed. H. Feigl and G. Maxwell. Vol. 3, *Minnesota Studies in the Philosophy of Science*. Minneapolis: University of Minnesota Press, 1962.
- . "Problems of Empiricism." In *Beyond the Edge of Certainty*. Ed. R. G. Colodny. Englewood Cliffs, N.J.: Prentice-Hall, 1965.
- . *Against Method*. London: New Left Books, 1975.
- Hempel, Carl G. "Reduction: Ontological and Linguistic Issues." *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*. Ed. S. Morgenbesser et al. New York: St. Martin's Press, 1969.
- Hull, David. "Reduction in Genetics—Biology or Philosophy?" *Philosophy of Science* 39 (1972): 491–99.
- . "The Reduction of Mendelian to Molecular Genetics." Chap. 1, *Philosophy of Biological Science*. Englewood Cliffs, N.J.: Prentice-Hall, 1974.

Kemeny John G., and Paul Oppenheim. "On Reduction." *Philosophical Studies* 7 (1956): 6–17.

Kitcher, Philip. "Explanatory Unification and the Causal Structure of the World." In *Scientific Explanation*. Ed. P. Kitcher and W. C. Salmon. Vol. 13, *Minnesota Studies in the Philosophy of Science*. Minneapolis: University of Minnesota Press, 1989.

———. "Darwin's Achievement." Chap. 2, *The Advancement of Science*. Oxford: Oxford University Press, 1993.

Kuhn, Thomas. "The Nature and Necessity of Scientific Revolutions." Chap. 9, *The Structure of Scientific Revolutions*, 3d ed. Chicago: University of Chicago Press, 1996.

Nagel, Ernest. "The Meaning of Reduction in the Natural Sciences." In *Science and Civilization*. Ed. R. C. Stauffer. Madison: University of Wisconsin Press, 1949. Reprinted in *Philosophy of Science*. Ed. A. Danto and S. Morgenbesser. Cleveland: Meridian Books, 1960.

———. "The Reduction of Theories." Chap. 11, *The Structure of Science*. New York: Harcourt, Brace and World, 1961.

Nye, Mary Jo. *Molecular Reality*. London: Macdonald, 1972.

Quine, W. V. "Ontological Reduction in the World of Numbers." *Journal of Philosophy* 61 (1964): 209–16.

Rosenberg, Alexander. "Reductionism and the Temptation of Provincialism." Chap. 4, *The Structure of Biological Science*. Cambridge: Cambridge University Press, 1985.

———. "From Reductionism to Instrumentalism?" In *What the Philosophy of Biology Is*. Ed. M. Ruse. Dordrecht, Netherlands: Kluwer Academic Publishers, 1989.

———. "Whatever Happened to Reductionism, and Why?" and "Reductionism and Explanation in Molecular Biology." Chaps. 2 and 3, *Instrumental Biology: or the Disunity of Science*. Chicago: University of Chicago Press, 1994.

Schaffner, Kenneth F. "Approaches to Reduction." *Philosophy of Science* 34 (1967): 137–47.

———. "The Watson-Crick Model and Reductionism." *British Journal for the Philosophy of Science* 20 (1969): 325–48.

Shapere, Dudley. "Meaning and Scientific Change." In *Mind and Cosmos*. Ed. R. G. Colodny. Pittsburgh, Pa.: University of Pittsburgh Press, 1966.

Sklar, Lawrence. "Types of Intertheoretic Reduction." *British Journal for the Philosophy of Science* 18 (1967): 109–24.

Spector, Marshall. *Concepts of Reduction in Physical Science*. Philadelphia, Pa.: Temple University Press, 1978.

Waters, C. Kenneth. "Why the Antireductionist Consensus Won't Survive the Case of Classical Mendelian Genetics." In *PSA 1990*. Vol. 1. Ed. A. Fine, M. Forbes, and L. Wessels. East Lansing, Mich.: Philosophy of Science Association, 1990.

———. "Genes Made Molecular." *Philosophy of Science* 61 (1994): 163–85.

Wimsatt, William. "Reduction and Reductionism." In *Current Issues in the Philosophy of Science*. Ed. P. D. Asquith and H. E. Kyburg, Jr. East Lansing, Mich.: Philosophy of Science Association, 1978.

CHAPTER 9 | EMPIRICISM AND SCIENTIFIC REALISM

Achinstein, Peter, and Stephen F. Barker, eds. *The Legacy of Logical Positivism*. Baltimore, Md.: Johns Hopkins Press, 1969.

Bourgeois, W. "On Rejecting Foss's Image of van Fraassen." *Philosophy of Science* 54 (1987): 303–8.

Boyd, Richard N. "The Current Status of Scientific Realism." In *Scientific Realism*. Ed. Jarrett Leplin. Berkeley: University of California Press, 1984.

Brown, James Robert. "The Miracle of Science." *Philosophical Quarterly* 32 (1982): 232–44.

Carrier, Martin. "What Is Wrong with the Miracle Argument?" *Studies in History and Philosophy of Science* 22 (1991): 23–36.

Cartwright, Nancy. *How the Laws of Physics Lie*. Oxford: Clarendon Press, 1983.

Churchland, Paul M., and Clifford A. Hooker, eds. *Images of Science: Essays on Realism and Empiricism*. Chicago: University of Chicago Press, 1985.

Clendinnen, F. John. "Realism and the Underdetermination of Theory." *Synthese* 81 (1989): 63–90.

Cushing, James T. *Quantum Mechanics: Historical Contingency and the Copenhagen Hegemony*. Chicago: University of Chicago Press, 1994.

Fine, Arthur. "And Not Anti-Realism Either." *Nous* 18 (1984): 51–65. Reprinted in *The Shaky Game: Einstein, Realism, and the Quantum Theory*. 2d ed. Chicago: University of Chicago Press, 1996.

———. "Unnatural Attitudes: Realist and Instrumentalist Attachments to Science." *Mind* 95 (1986): 149–79.

———. "Piecemeal Realism." *Philosophical Studies* 61 (1991): 79–96.

———. *The Shaky Game: Einstein, Realism, and the Quantum Theory*. 2d ed. Chicago: University of Chicago Press, 1996.

Foss, Jeff. "On Accepting van Fraassen's Image of Science." *Philosophy of Science* 51 (1984): 79–92.

———. "On Saving the Phenomena and Mice: A Reply to Bourgeois Concerning van Fraassen's Image of Science." *Philosophy of Science* 58 (1987): 278–87.

Gutting, Gary. "Scientific Realism versus Constructive Empiricism: A Dialogue." *The Monist* 65 (1982): 336–49. Reprinted in *Images of Science: Essays on Realism and Empiricism*. Ed. Paul M. Churchland and Clifford A. Hooker. Chicago: University of Chicago Press, 1985.

Hacking, Ian. "Do We See through a Microscope?" *Pacific Philosophical Quarterly* 62 (1981): 305–22. Reprinted in *Images of Science: Essays on Realism and Empiricism*. Ed. Paul M. Churchland and Clifford A. Hooker. Chicago: University of Chicago Press, 1985.

———. *Representing and Intervening*. Cambridge: Cambridge University Press, 1983.

Hardin, Clyde L., and Alexander Rosenberg. "In Defense of Convergent Realism." *Philosophy of Science* 49 (1982): 604–15.

Hausman, Daniel M. "Constructive Empiricism Contested." *Pacific Philosophical Quarterly* 63 (1982): 21–28.

Hempel, Carl G. "The Theoretician's Dilemma." *Aspects of Scientific Explanation*. New York: Free Press, 1965.

Hobbs, Jesse. "A Limited Defense of the Pessimistic Induction." *British Journal for the Philosophy of Science* 45 (1994): 171–91.

Hooker, Clifford A. "Surface Dazzle, Ghostly Depths: An Exposition and Critical Evaluation of van Fraassen's Vindication of Empiricism against Realism." In *Images of Science: Essays on Realism and Empiricism*. Ed. Paul M. Churchland and Clifford A. Hooker. Chicago: University of Chicago Press, 1985.

Kukla, André. "Scientific Realism, Scientific Practice, and the Natural Ontological Attitude." *British Journal for the Philosophy of Science* 45 (1994): 955–75.

———. "The Two Antirealisms of Bas van Fraassen." *Studies in History and Philosophy of Science* 26 (1995): 431–54.

———. "The Theory-Observation Distinction." *Philosophical Review* 105 (1996): 173–230.

Leplin, Jarrett, ed. *Scientific Realism*. Berkeley: University of California Press, 1984.

Laudan, Larry. *Science and Hypothesis*. Dordrecht, Netherlands: D. Reidel, 1981.

McMichael, Alan. "Van Fraassen's Instrumentalism." *British Journal for the Philosophy of Science* 36 (1985): 257–72.

McMullin, Ernan. "Comment: Selective Anti-Realism." *Philosophical Studies* 61 (1991): 97–108.

Newton-Smith, W. H. *The Rationality of Science*. London: Routledge and Kegan Paul, 1981.

O'Connor, D. J. *The Correspondence Theory of Truth*. London: Hutchinson and Co., 1975.

Oddie, Graham. "Truthlikeness." In *The Encyclopedia of Philosophy Supplement*. Ed. D. M. Borchert. New York: Simon and Schuster Macmillan, 1996.

O'Leary-Hawthorne, John. "What Does van Fraassen's Critique of Scientific Realism Show?" *The Monist* 77 (1994): 128–45.

Putnam, Hilary. "The Meaning of Meaning." In *Language, Mind, and Knowledge*. Vol. 7, *Minnesota Studies in the Philosophy of Science*. Ed. Keith Gunderson. Minneapolis: University of Minnesota Press, 1975. Reprinted in *Mind, Language, and Reality: Philosophical Papers*. Vol. 2. Cambridge: Cambridge University Press, 1975.

———. "What Is Mathematical Truth?" In *Mathematics, Matter, and Method: Philosophical Papers*. Vol. 1. Cambridge: Cambridge University Press, 1975.

Reiner, Richard, and Robert Pierson. "Hacking's Experimental Realism: An Untenable Middle Ground." *Philosophy of Science* 62 (1995): 60–69.

Salmon, Wesley C. *Scientific Explanation and the Causal Structure of the World*. Princeton, N.J.: Princeton University Press, 1984.

Scheffler, Israel. *The Anatomy of Inquiry*. New York: Knopf, 1963.

Sellars, Wilfrid. *Science, Perception and Reality*. New York: Humanities Press, 1962.

———. "Is Scientific Realism Tenable?" In *PSA 1976*. Vol. 2. Ed. F. Suppe and P. D. Asquith. East Lansing, Mich.: Philosophy of Science Association, 1977.

Smart, J. J. C. *Between Science and Philosophy*. New York: Random House, 1968.

Suppe, Frederick. "The Search for Philosophical Understanding of Scientific Theories." In *The Structure of Scientific Theories*, 2d ed. Ed. F. Suppe. Urbana: University of Illinois Press, 1977.

Van Fraassen, Bas C. "Wilfrid Sellars on Scientific Realism." *Dialogue* 14 (1975): 606–16.

———. *The Scientific Image*. Oxford: Clarendon Press, 1980.

———. "Empiricism in Philosophy of Science." In *Images of Science: Essays on Realism and Empiricism*. Ed. Paul M. Churchland and Clifford A. Hooker. Chicago: University of Chicago Press, 1985.

NAME INDEX

- Abraham, Max, 316
 Achinstein, Peter, 405n28, 406n30,
 411, 469, 470, 478n9, n14, 479n19,
 n21, 492n4, 533–38, 543, 546n23,
 547n39, 667, 721–22, 726, 734,
 742n8, 744n16, n17, 784, 785, 786,
 800, 843n6, 970n36, 997n8, 1298
 Adams, John Couch, 304, 461
 Addelson, Kathryn Pine, 187n10
 Adler, Alfred, 4, 5, 8, 318n2. *See also*
 (in subject index) Psychology,
 individual
 Alberch, P., 1002n39, n40
 Aldrin, Edwin (“Buzz”), 452n
 Alexander, R. McNeill, 866
 Alexander the Great, 28
 Ampère, André Marie, 22, 363, 398
 Ångström, Anders Jonas, 512
 Aquinas, Saint Thomas, 1099, 1104
 Arago, François, 166, 263, 265, 278,
 305, 528, 529
 Archimedes, 772
 Ariew, Roger, 318n4
 Aristotle, 131, 179, 197, 281, 297,
 341, 378, 409, 679, 726, 727, 727n,
 733, 739, 742n9, 788, 932,
 1086n16, 1212, 1296. *See also* (in
 subject index) Aristotelianism/
 Aristotelians; Logic, Aristotelian;
 Physics, Aristotelian
 Armstrong, D. M., 807n2, 858, 860–
 61, 862, 885, 900n4, 1310
 Armstrong, Neil, 452n
 Asquith, P. D., 1086n17, 1204n3,
 1206n15, 1207n26
 Augustine, Saint, 1084, 1099
 Austin, J. L., 1170n
 Avenarius, R., 455
 Ayer, A. J., 30, 367, 806, 807n2,
 843n6, 850, 879–85, 898, 1286n3,
 1296, 1297, 1301, 1304
 Bacon, Francis, 150, 161, 170, 264,
 305, 355, 359, 431n5, 497, 1156
 Bain, Alexander, 1166, 1167
 Balmer, John Jacob, 450–51, 452,
 512, 516, 965, 967n2, 1087n27
 Bambrough, Renford, 741n1
 Bamford, Greg, 81n14
 Barash, David, 177
 Barker, Peter, 1206n16
 Barker, Stephen F., 405n28, 406n30,
 459n9, 704n6, n7, 718n3, 970n36,
 997n8
 Barnes, Barry, 393
 Bartholin, Erasmus, 1168n10
 Bayes, Thomas, 464, 528, 549–50,
 551, 553–58, 559, 565, 568, 569,
 570, 574, 577, 578–79, 581n16,
 588, 600, 601, 623n10, 628–34,
 639, 640, 642, 648, 656, 667, 668,
 797, 1292
 Beadle, George, 980
 Beatty, John, 188n29
 Bellarmine, Cardinal Robert, 1129
 Bentham, Jeremy, 293
 Benveniste, Jacques, 177
 Bergmann, Gustav, 1055, 1062n5,
 1229–30
 Bergson, Henri, 430, 430n, 507
 Berkeley, George, 366–67, 1091,
 1111n3, 1166, 1215, 1216, 1221,
 1222, 1286n3, 1304
 Bernard, Claude, 257, 259, 279, 313,
 355, 358
 Bertolet, Rod J., 406n29
 Bhattacharya, Gouri, 1184n10
 Bigelow, John, 807n2
 Biot, Jean-Baptiste, 263, 265, 278,
 309, 359, 365, 401n4, 528
 Birkhoff, Garrett, 380
 Black, Joseph, 1121
 Black, Max, 539, 744n19, 825n12
 Bloor, David, 156, 202–203, 255,
 256, 321, 342, 344–46, 348n4,
 353n59, 383, 389, 393–97, 400–
 401, 407n37, 1308, 1309
 Boas, Maria, 101n5
 Bode, Johann Elert, 1101n
 Boerhaave, Herman, 1121
 Bohm, David, 924n, 925, 928, 1196,
 1240
 Bohr, Niels, 116n, 124, 229, 452,
 516, 919, 924n, 951, 964, 1052,
 1118, 1119, 1156, 1186, 1195,
 1196, 1198, 1207n26, 1269
 Bok, Bart J., 29, 31
 Boltzmann, Ludwig, 920, 1045n12,
 1064
 Born, Max, 924n, 927, 1052n, 1195

- Bose, Satyendranath, 1168n8
 Bourgeois, Warren, 1286n10
 Boyd, Richard, 321, 1083–84,
 1085n4, n5, 1114, 1115, 1117–18,
 1122, 1123, 1129, 1131, 1132n1,
 n3, 1133n7, 1140–41, 1142, 1177,
 1181, 1187, 1205n7
 Boyle, Robert, 113, 130, 652, 688,
 726, 741, 805, 973, 987, 997n7,
 1036, 1066, 1104
 Bradley, F. H., 862
 Brahe, Tycho, 13, 17, 222–23, 269,
 270, 463, 464, 567–68
 Braithwaite, R. B., 45n5, 101n4, 551,
 581n18, 807n2, 816, 836–37,
 843n6, 846, 848, 851, 862, 1294
 Brandon, Robert N., 80n2
 Brewster, David, 572
 Bridgman, Percy W., 1303
 Bridges, C. B., 989
 Brill, A.A., 694n5
 Broad, C. D. 186n2, 497, 694n2,
 817–18, 882
 Brodbeck, May, 945n7, 1062n4
 Brody, Baruch, 721, 728, 785, 800,
 837, 844n10
 Broglie, Louis de, 564, 928, 1057,
 1196
 Bromberger, Sylvain, 726, 843n6,
 1000n22, 1107
 Brouwer, L. E. J., 317, 379, 1187
 Brown, Frank, 35n4
 Brown, James Robert, 348n7, 1050,
 1253–61, 1281–82, 1288n25
 Brown, Robert, 1013
 Brush, Stephen G., 465–66, 477n6,
 481, 482, 486, 488–89, 490,
 491n1, 514, 528–31, 536, 546n30,
 561, 582n32
 Buchdahl, Gerd, 187n13, 396,
 408n43
 Buchwald, Jed Z., 492n4, 547n39
 Bunge, M., 944n2
 Burian, Richard M., 546n30, 1120
 Butts, Robert, 742n7

 Cajori, F., 82n20, 151n, 1112n12,
 1113n13
 Campbell, Norman, 1228
 Cantor, Georg, 1187, 1268
 Cardwell, D. S. L., 91n
 Cargile, James, 849
 Carnap, Rudolf, 95n, 139, 282, 283,
 289, 291, 294, 295, 298, 299,
 299n7, 300n13, 312, 366, 367, 372,
 373, 402n10, 403n14, 409, 431n5,
 458, 464, 469, 479n21, 482, 492n2,
 510, 516, 524, 534, 541, 546n28,
 551, 553, 557n, 580n2, 585, 586,
 588, 593, 613, 616–17, 621n3,
 623n12, 640, 665, 676, 689, 703,
 711, 718n3, 767–69, 770, 782–83,
 788, 789, 799, 807n2, 844n10, 931,
 1070–71, 1228, 1285n1, 1286n2,
 1294, 1298, 1300
 Carpenter, C. R., 204
 Carroll, John, 807n2
 Cartwright, Nancy, 806, 874, 876,
 896–97, 899, 901n11, n12, 1146,
 1151n15, 1167n3, 1168n15, 1171,
 1184n3, 1205n6, 1205n6, 1207n27,
 1288n24, n26
 Cassini, Jean Dominique, 341,
 353n53
 Castañeda, H.-N., 1086n12
 Causey, Robert, 997n8
 Cavendish, Henry, 97n, 565, 1121
 Ceci, Stephen, 187n8
 Chalmers, Alan, 901n12
 Charles, Jacques, 113, 726, 741, 973,
 987, 997n7, 1036
 Chastain, Charles, 1199
 Chomsky, Noam, 412–13, 419, 420,
 496–97
 Christensen, David, 621n3
 Church, Alonzo, 1142
 Cicero, 28
 Cioffi, Frank, 35n6
 Clagett, Marshall, 352n48
 Clairault, Alexis, 71, 81n13
 Clarke, Samuel, 1113n13
 Clarke, Steve, 901n12
 Clausius, Rudolph, 920, 935
 Clavius, Christopher, 1100
 Coffa, J. Albert, 804n16
 Cohen, I. B. 101n7, 949n39
 Cohen, L. Jonathan, 406n31
 Cohen, R. S., 82n21, 137n26,
 404n21, 967n2, 996n3, n4,
 1206n12
 Cole, Jonathan R., 187n8
 Cole, Stephen, 187n8
 Collingwood, R. G., 394
 Collins, Harry M., 79n1, 321, 348n5
 Collins, Wilkie, 922
 Colodny, R.G., 118n4, 581n20,
 606n10, 724n7, 921n12, 946n16
 Comte, Auguste, 1166, 1304

- Conant, James B., 12n, 100n3, 458n5
 Copernicus, Nicolaus, 13, 17, 26, 68,
 84, 86, 103–105, 109, 111–12,
 123, 132–35, 137n12, 229–30,
 232–33, 234, 340, 459n6, 586,
 599, 1078, 1188n, 1307. *See also*
 (in subject index) Revolution, (Co-
 pernicus; Astronomy, Copernican
 Cotes, Roger, 1113n13
 Coulomb, C. A., 868–75
 Cox, R. T., 621n3
 Cracraft, J., 1002n38
 Creary, Lewis, 901n11
 Crick, Francis, 903, 971, 975, 985,
 988, 996, 996n2, 1028, 1035
 Crookes, William, 87, 490
 Curd, Martin, 251n3, 544n7, 545n11,
 n14
 Currie, Greg, 1113n17
 Cushing, James T., 81n18, 924n,
 1288n30

 Dalton, John, 97, 143–44, 165, 166,
 238, 586, 1118, 1119, 1156
 Daniken, Erich von, 1, 74
 Danto, A. C., 945n8
 Darden, Lindley, 996n4
 Darrow, Clarence, 52
 Darwin, Charles, 43, 45, 51, 58, 65–
 66, 106, 151, 194, 225, 251n1, 297,
 697, 775, 1138, 1307. *See also* (in
 subject index) Darwinism
 Davidson, Donald, 744n21, 850
 Davidson, Eric H., 999n17
 Davisson, Clinton J., 564
 Dawkins, Richard, 1002n38
 Delaney, C. F., 81n18
 Democritus, 905
 Delisle, J. N., 533
 Derrida, Jacques, 321
 Descartes, René, 34, 82n20, 96, 124,
 161, 225, 341, 359, 392, 415, 416,
 417, 431n5, 940
 Dewey, John, 152, 685, 1157, 1166n
 Dicke, Robert H., 449n
 Dobzhansky, Theodosius, 973, 998n9
 Doell, Ruth, 194–95
 Donagan, Alan, 677n1
 Doolittle, Eric, 601, 606n18
 Doppelt, Gerald, 162, 165–67,
 169n31, n38, 238
 Dorling, Jon, 586, 606n17, 672n10
 Dray, William H., 677n1, 694n2,
 1151n17

 Dretske, Fred I., 806, 807n2, 826n,
 828n, 858, 861, 862, 880, 883,
 884–89, 896, 898–99, 900n1, n4,
 1310
 Driesch, Hans, 676, 679–82, 768–69,
 770
 Dugas, René, 101n6
 Duhem, Pierre, 30, 82n20, 255–56,
 279n10, n11, 300n17, 302n17,
 318n2, n3, n4, 320, 327, 330, 342,
 344, 350n19, 355–65, 374–78,
 379, 381–82, 383, 384, 388, 397–
 98, 399–400, 401n1, n2, 402n5,
 nn7–9, 431n5, 519, 582n27, 639,
 767–68, 802n3, 803n5, 928,
 947n25, 1016, 1071, 1100, 1101,
 1104, 1164, 1188, 1291, 1294,
 1295, 1296, 1298, 1299, 1309. *See*
also (in subject index) Duhem the-
 sis; Duhem-Quine thesis
 Dummett, Michael, 855, 1066, 1067,
 1068, 1082–83, 1087n25
 Dunn, L. C., 998n13
 Durrant, Bob, 1113n17

 Earman, John, 79n1, 138n40, 402n7,
 478n18, 547n31, 550n1, 621n3,
 623n8, n10, n13, 658, 673n23,
 n26, 1194, 1206n14, n15, 1295
 Eddington, Arthur E., 4, 6, 63, 79n1,
 529, 905, 1222n
 Ede, Donald, 1002n36, n39
 Edey, M., 47n21
 Edge, David, 393
 Edgley, Roy, 925
 Eells, Ellery, 621n3
 Egidi, R., 348n7
 Ehrenfest, Paul, 960, 962
 Ehrenhaft, Felix, 178, 938,
 947n27
 Einstein, Albert, 4–7, 9n1, 13, 22,
 24–25, 29, 63, 68, 79n1, 91–94,
 115–16, 116n, 124, 158, 164, 177,
 184, 213, 297, 308–9, 310, 317,
 318n2, 370, 383, 430, 431n6, 449n,
 452, 465–66, 477n7, 507, 512,
 513, 529, 530–31, 558–59, 586,
 599, 601, 606n20, 656, 658–59,
 747n, 923, 931, 936, 946n22,
 947n29, 952, 962, 967n5, 1013,
 1019–20, 1045n10, 1046n18, n20,
 1111n4, 1186, 1194–96, 1198,
 1206n15, 1222, 1269, 1303. *See*
also (in subject index) Dynamics,

- Einstein, Albert (*cont.*)
 Einsteinian vs. Newtonian; Gravitation, theory of, Einstein's
 Eldredge, Niles, 1002n38
 Eliot, T. S., 1217n
 Ellis, Brian, 1065
 Empedocles, 8
 Enç, Berent, 997n8
 Euclid, 264, 315, 368, 370
 Evans, E., 458n4
 Everitt, C. W. F., 1167n1
- Faraday, Michael, 492n6, 547n39, 923
 Fausto-Sterling, Anne, 196, 198
 Feigl, Herbert, 551, 580n2, 694n2, n3, 704n6, 802n2, 945n6, n7, 968n9, 1046n17, 1062n2, n4, n6, 1301
 Feinberg, Gerald, 187n7
 Fetzer, James H., 582n26
 Feyerabend, Paul K., 15n, 27n, 50, 92n, 126, 172–73, 223, 240, 321, 363, 704n6, 904, 915–19, 924n, 944n2, 946n16, n21, 947n29, 949nn42–44, 950, 955, 956, 960, 961, 1011–22, 1025, 1026, 1028, 1039–41, 1044n7, 1045n8, n11, 1060, 1061, 1062n2, 1071, 1172n, 1293, 1299, 1305, 1309
 Feynman, Richard, 807n2, 866, 867–68
 Fichte, Johann Gottlieb, 678
 Field, Hartry, 621n3, 1085, 1214
 Fine, Arthur, 116n, 132, 255, 348n7, 406n35, 901n11, 1047n25, 1051, 1206n12, n16, 1207n22, n25, 1209–25, 1267–77, 1283–85, 1288n33, 1289n35, 1302
 Fine, Kit, 1085n8
 Finetti, Bruno de, 621n3, 623n8, 635
 Fishburn, Peter, 624n16
 Fitzgerald, G. F., 13n, 178, 571, 1019
 Fizeau, H. L., 108
 Fleischmann, Martin, 468–69, 469n, 472
 Fodor, Jerry, 997n6
 Forbes, Micky, 546n30, 147n25
 Forster, Malcolm, 652
 Foss, Jeff, 1242, 1286n10
 Foucault, J. L., 108, 108n, 263, 265, 278, 305–6, 309–10, 359–60, 362, 401n4, 572
 Foucault, Michel, 140, 144, 155
- Frank, Philipp, 402n8
 Franklin, Allan, 621n3, 1163n, 1183
 Franklin, Benjamin, 97
 Franklin, Melissa, 1168n7
 Franklin, Rosalind, 187n7
 Franz, Rudolph, 957, 958, 959, 1026
 Frederick, Robert E., 1160n
 Frege, Gottlob, 281, 293, 311, 317, 368–71, 378, 399, 585, 830, 893, 1291, 1301, 1308
 French, Peter A., 544n4
 Fresnel, Augustin, 265, 278, 309, 359, 362, 365, 401n4, 528, 529, 1121, 1148
 Freud, Sigmund, 1, 4–6, 8, 9n3, 563, 582n24, 584, 692–93, 773. *See also* (in subject index) Psychoanalysis
 Freundlich, Erwin F., 606n20
 Fricke, Martin, 1113n17
 Fried, Barbara, 207n2
 Friedman, Michael, 745n25, 1114
 Fumerton, Richard, 543n2
 Fürth, R., 947n28, n29
- Galen, 197
 Galileo, 10n4, 51, 105, 112, 123–24, 127, 132, 135, 226, 234, 448n, 455, 473, 479n20, 517, 567, 574–75, 686, 687, 688, 905, 908, 909, 923, 929, 932, 944n2, 945n3, 968n14, 972, 997n7, 1007, 1016, 1038, 1040, 1183, 1298
 Galison, Peter, 189n35, 406n35, 1168n13
 Garber, Daniel, 606n17, 673n26
 Gardiner, Patrick, 677n1, 744n19
 Gardner, Martin, 2n1, 13n, 449n, 562, 1160n
 Gardner, Michael, 477n2, 138n40, 1167
 Garfinkel, Alan, 1000n20
 Gauquelin, Michel, 30–31, 35n4
 George, Rolf A., 294n
 Georgioui, G. P., 1001n30
 Gerlach, Walter, 1156
 Gibbins, Peter, 404n20, n23
 Giere, Ronald N., 477n3, 528, 581n15, 621n3, 761, 807n1, 1204n3, 1207n26, 1309
 Gillies, Donald, 255–56, 303n, 366, 374, 375–82, 399–400, 401n3, 647
 Gillispie, C. C., 101n11
 Gingerich, Owen, 528, 530
 Gish, Duane T., 41–42, 57, 59–60

- Glazer, Sarah, 187n8
 Glymour, Clark, 79n1, 83, 85n1, 204,
 423, 469, 471, 482, 550, 605n1,
 623n10, 634–35, 646–59, 661,
 664, 668–70, 671n6, 672n19, n20,
 673n22, n24, n26, 742n7, 997n6,
 n8, 1085n6, 1122, 1206n14
 Gödel, Kurt, 1188
 Goldbach, Christian, 379, 811–12,
 856
 Goleman, Daniel, 187n8
 Good, I. J., 621n3, 623n9, 673n31
 Goodling, David, 411n2, 546n29
 Goodman, Nelson, 156, 336–38, 383,
 389–91, 400–401, 405n28,
 406n30, 421, 422, 503, 544n8, 660,
 819–20, 843n6, 849
 Goosens, William, 741n4
 Goudge, Thomas A., 1151n16
 Goudsmit, Samuel, 1156
 Gould, Stephen Jay, 49, 177, 188n32,
 1002n38, n39
 Greene, H. C., 279n1
 Greeno, J., 742n7
 Greenstreet, W. J., 303n, 813n
 Grene, Majorie, 174
 Grice, H. P., 403n15
 Grimaldi, Francesco Maria, 274
 Groenewold, H. J., 945n4
 Grosser, Morton, 81n14
 Grossmann, Marcel, 606n18
 Grove, Jack W., 2n1
 Grünbaum, Adolf, 328, 349n13, 350,
 n19, 352n49, 405n27, 444n3,
 545n17, 575, 581n9, 582n24, n31,
 704n6, n7
 Guerlac, Henry, 18n5
 Gunderson, Keith, 1173n, 1204n4
 Goosens, William K., 996n3
 Gutting, Gary M., 81n18, 348n7

 Haack, Susan, 544n5
 Habermas, Jürgen, 183, 188n30, 245,
 252n13
 Hacking, Ian, 606n17, 621n3, 1051,
 1150, 1167n1, n2, 1168n6, n18,
 1169–83, 1253, 1261–67, 1282–
 83, 1288n26
 Hadamard, Jacques, 279n11
 Hafner, E. M., 18n5
 Hahn, Hans, 367, 1301
 Halley, Edmund, 25
 Halstead, Beverly, 80n8
 Hannaway, Owen, 546n23

 Hanson, N. R., 15n, 947n26, 1062n2,
 1291, 1309
 Harari, Haim, 582n23
 Haraway, Donna, 203
 Hardin, Clyde L., 1251
 Harding, Sandra, 188n33, 200, 205,
 214, 246–49, 255, 349n12, n14,
 n16, 350n18, n19, 405n27
 Hardy, G. H., 39
 Harman, Gilbert, 462, 498–99, 837,
 1075, 1086n12, 1299
 Harper, W., 606n15
 Harré, Rom, 547n40
 Harris, A., 431n6
 Hartley, David, 1121, 1146
 Harvey, E., 208n15
 Harvey, William, 1183
 Hausman, Alan, 403n13
 Healey, Richard, 745n25
 Hegel, G. W. F., 199, 678, 1220n
 Heisenberg, Werner, 924, 947n26,
 1052, 1057, 1195, 1269
 Helmholtz, Hermann von, 151, 311
 Hempel, Carl G., 45n5, 128, 139,
 186n3, 252n12, 300n15, 351n25,
 367, 370, 402n8, 403n13, 410, 420,
 422, 427n, 448n, 449n, 469, 482,
 507, 510–24, 540–42, 544n10,
 546n24, 551, 579, 627, 640, 659–
 60, 669, 675–77, 677n1, 694n2,
 n3, 699n, 717n2, 718n7, 720–33,
 736, 742n5, n9, 15, 749–52, 762–
 63, 765n2, 767–93, 794, 796, 799–
 801, 803n12, 807n2, 826, 873n,
 843n1, n6, 844n8, 903, 945n7, 962,
 n32, 997n5, n6, 1148, 1149, 1150,
 1185n18, 1228, 1260, 1282,
 1285n1, 1286n2, 1288n25, 1294,
 1295, 1296, 1299, 1300, 1301,
 1303, 1306, 1307, 1309
 Henifin, Mary Sue, 207n2
 Herschel, William, 71, 130, 161
 Hertz, Heinrich, 151, 490–91, 492n6,
 537–38, 547n38, n39, 1064
 Hesse, Mary B., 156, 188n19, 255,
 321, 331, 342–44, 350n22, 383,
 389, 392, 395–96, 400–401,
 407n36, 591, 592, 593, 595–96,
 621n3, 648, 651, 652–53, 672n12,
 673n22, 846, 849, 962
 Hilbert, David, 948n30, 1187–89,
 1268
 Hill, E. L., 945n6, 946n22
 Hilpinin, Risto, 347n3

- Hintikka, Jaakko, 582n25, 586, 621n3, 622n4, 742n7
- Hobbes, Thomas, 34, 808
- Hoffmann, Banesh, 1206n14
- Hollinger, Robert, 252n13
- Holton, Gerald, 140, 188n21, 458n2, n3, 561, 582n32, 947n27, 1205n7, 1206n12
- Homer, 8, 298
- Hooke, Robert, 457, 726, 770, 882, 883, 947n26, 1297
- Hooker, C. A., 606n15, 1085n5, 1086n22, 1113n17
- Horwich, Paul, 406n32, 527, 550, 606n16, 621n2, 622n6, 624n15, 650, 651, 655, 659–67, 670–71, 671n4, 674n34, 1204n1, 1207n27
- Houdini, Harry, 1182
- Howard, Don, 318n2
- Howson, Colin, 12n, 80n11, 406n32, 407n40, 478n15, n17, 611, 621n3, 657, 673n26
- Hoyle, Fred, 44
- Hubbard, L. Ron, 562
- Hubbard, Ruth, 194–95, 198
- Hubble, Edwin P., 882
- Hull, David L., 46n8, 189n35, 546n30, 973, 996n3, 1305
- Hume, David, 21, 130, 150, 280, 283, 296, 326, 368, 383, 390, 405n25, 410, 415–18, 420, 421, 423, 427, 496, 497–501, 505–6, 543n3, 738, 806, 811–12, 814, 848, 856, 871, 879–80, 888, 891, 897, 898, 900n1, n5, 1059, 1060, 1129, 1176–77, 1186, 1194, 1202, 1203, 1264, 1296, 1306, 1308
- Humphreys, Paul, 581n16, 804n17
- Hutton, James, 1121
- Huygens, Christian, 96, 130, 265, 305, 571
- Jaffe, Bernard, 188n22
- Jammer, Max, 960, 970n35
- Jardine, Nicholas, 138n40
- Jaspers, Karl, 948n37
- Jaynes, E. T., 621n3
- Jeffrey, Richard C., 546n23, 603, 605n1, n5, 606n17, 621n3, 624n16, 657, 671n6, 673n26, 742n7, 751–52, 792, 1085n6
- Jeffreys, Harold, 594–95, 621n3, 651–52
- Jensen, A., 177
- Jerome, Lawrence, 29, 31
- Jobe, Evan, 742n8
- Johanson, Donald, 47n21
- Johnson, Richard, 1184n10
- Jones, Roger, 1207n17
- Joyce, James, 405n28
- Julius Caesar, 28
- Kamin, Leon, 188n32
- Kamp, Hans, 1085n8
- Kant, Immanuel, 40–41, 115, 177, 252n9, 280, 311, 312, 368–71, 428, 429, 506, 1160, 1219–21, 1291
- Kaplan, Mark, 621n3
- Kaufmann, Walter, 24
- Keller, Evelyn Fox, 253
- Kelvin, Lord (William Thomson), 87, 87n, 586, 935
- Kepler, Johannes, 10n4, 17, 21, 28, 33, 39–40, 104–6, 112, 130, 134, 220–23, 226, 234, 266–70, 297, 303, 363, 364, 382, 398, 463, 465, 517, 574–75, 586, 595, 598–99, 655, 656, 675, 686, 687, 688, 771, 908, 909, 910, 929, 968n14, 991, 1007, 1016, 1038, 1040, 1046n15, 1188n
- Keynes, John Maynard, 431n3, 459n10, 491, 521, 621n3
- Kiefer, H., 968n14
- Kim, Jaegwon, 730, 730n, 788
- Kimbrough, Steven Orla, 996n3
- King, John, 1208n29
- Kirk, Bill, 1168n7
- Kirchhoff, Gustav, 678
- Kitcher, Philip, 136n5, 677n1, 745n25, 904, 981n, 998n10, 1000n22, 1001n23, 1002n34, n35, 37, 1028–38, 1042–43, 1046n22, 1047n24, n27, 1305
- Klemke, E. D., 252n13
- Kline, A. David, 252n13, 901n11
- Kneale, W. C., 807n2, 811–12, 824n6, 843n3, 852
- Koestler, Arthur, 28
- Koplik, Henry, 702, 777–78
- Kordig, Carl, 251n4
- Kornberg, Arthur, 988, 1001n29
- Körner, Stephan, 744n19, 946n14, 947n25
- Kourany, Janet A., 251n3
- Kraft, Viktor, 431n5, 946n20
- Krajewski, W., 1127

- Kramers, H. A., 962
 Kripke, Saul, 806, 855, 857–58, 888,
 891–93, 898, 900n6, n7, n8, 1172,
 1299, 1306, 1308
 Kronecker, Leopold, 1187
 Kuhn, Thomas S., 2, 15n, 17n2,
 18n6, n8, n9, n16, n17, n20, 23,
 26, 32–34, 45n5, 50, 57, 63, 66–
 69, 71–73, 78, 83, 84, 85n1, n2,
 87n, 92n, 95n, 97n, 98n, 102,
 104n, 107n, 116n, 119–28, 130–
 36, 139–45, 149–50, 152–67, 172–
 73, 182, 188n20, 215–40, 242–44,
 247, 249–50, 251n6, 252n9, 255,
 256, 321, 338–40, 352n41, n45,
 383, 389, 391–93, 400–401,
 406n33, 413, 419, 458n5, 496–97,
 528, 551, 552–53, 555, 556, 562,
 569, 573, 574, 575–77, 578, 579–
 80, 580n8, 638, 640, 644, 645–46,
 668, 671n8, 904, 925, 947n26, 955,
 1000n22, 1019, 1025, 1034,
 1045n8, 1172, 1180, 1200,
 1208n30, 1293, 1294, 1299, 1300,
 1302, 1303, 1305, 1306, 1307, 1309
 Külpe, O., 431n3
 Kurtz, Paul, 29
 Kyburg, Henry E., Jr., 557n, 605n6,
 621n3, 623n8, 672n15, 803n13

 Lack, David, 47n17
 Lakatos, Imre, 2, 30, 32–33, 50, 63,
 69, 71–73, 78, 81n14, 85n2,
 118n3, 140, 144, 155, 321, 347n2,
 348n5, 349n16, 363, 364, 394, 434,
 547n34, 1295, 1297, 1307
 Landé, Alfred, 1195
 Langmuir, Irving J., 87n
 Laplace, Pierre, 263, 265, 359, 365,
 528, 586, 1121, 1292
 Latour, Bruno, 187n9, 188n14
 Laudan, Larry, 2, 54–56, 58, 60, 77,
 81n18, 82n21, n22, 84, 125–26,
 137n26, 132, 140, 145n, 168nn14–
 19, 187n4, n13, 202–3, 220, 224,
 228, 233, 235–39, 250, 256, 328,
 348n5, n7, 350n18, 352n40, 366,
 382–97, 400–401, 405n25, n26,
 n27, 406n33, n35, 407n41, 411n2,
 647, 918n, 1050, 1089, 1101,
 1111n10, 1145, 1146, 1147, 1148,
 1185n17, 1187, 1190, 1209–10,
 1246–52, 1253, 1255, 1259, 1259,
 1267, 1268, 1280–81, 1287n13,
 n16, 1292, 1295, 1297, 1306, 1309
 Laue, Max von, 316
 Lavoisier, Antoine, 12, 68, 86, 97, 98,
 104n, 114–15, 123, 137n12, 143–
 44, 145, 157, 231, 586, 1121
 Lawson, R. W., 13n
 Laymon, Ronald, 402n7, 1045n11
 Lea, C. H., 948n32
 Leavitt, Henrietta Swan, 445,
 446n
 Lee, Tsung Dao, 13, 449
 Leibniz, G. W., 225, 251n7, 280,
 282, 286, 1168n9
 Leplin, Jarrett, 138n34, 1147,
 1184n12
 Le Roy, Edouard, 271, 272, 306
 LeSage, G. L., 1121, 1146
 LeVerrier, U. J. J., 304, 461, 601
 Levi, Isaac, 621n3
 Levin, Michael, 1212–13
 Levins, Richard, 188n31
 Lewis, C. I., 299, 299nn6–7, 300n15,
 717n2
 Lewis, David, 647, 741n4, 765n9,
 852, 853, 855
 Lewontin, Richard, 45, 177, 188n31,
 n32, 998n13, 1002n38
 Liebig, Justus von, 431n5, n6
 Lindberg, David, 137n17
 Lindley, D. V., 624n16
 Lipton, Peter, 252n9, 410, 495–505,
 539, 542, 545n12, 724, 848, 1179,
 1299
 Lloyd, Elizabeth, 198–99
 Locke, Don, 187n12
 Locke, John, 293, 296, 893, 1059,
 1157, 1296
 Longino, Helen, 85, 188n24, 194,
 208n12, 212, 213–14, 239–45,
 247, 248, 250, 252n15, 255,
 547n35, 1293
 Lorentz, H. A., 13, 178, 571, 931,
 946n22, 1019–20, 1156
 Lowinger, Armand, 300n17
 Lucretius, 28
 Lugg, Andrew, 1134n14
 Lukaszewicz, Jan, 378
 Lycan, William, 1204n1
 Lyell, Charles, 56
 Lyon, Ardon, 720–21, 785
 Lysenko, T. D., 200, 252n14

 MacCullagh, James, 359
 Mace, C. A., 3n

- Mach, Ernst, 151, 431n6, 455, 519, 520, 678, 767, 768, 803n5, 807n1, 927, 930, 941, 946n23, 1166, 1194, 1204, 1227, 1304
- Machamer, Peter K., 348n7, 403n13, 901n11
- Mackie, J. L., 843n6, 851, 1107–8
- Madden, E. H., 459n11, 557n
- Maddox, John, 177
- Magie, W. F., 448n, 458n1
- Maher, Patrick, 621n3
- Mannheim, Karl, 918, 918n
- Martin, R. M., 300n14
- Martin, R. N. D., 310
- Marx, Karl, 4–6, 8, 252n13. *See also* (in subject index) Marxism
- Mattheson, Carl A., 901n11
- Maugh, T. H., II, 1001n31
- Mauil, Nancy, 996n4
- Maxwell, James Clerk, 98, 98n, 99, 229, 265, 310, 311, 547n38, 586, 681, 805, 865, 915, 1045n12, 1092, 1121, 1297
- Maxwell, Grover, 411n3, 704n6, 802n2, 945n6, 947n29, 968n9, 1046n17, 1050, 1062n2, n6, 1071–74, 1229–36, 1277–79, 1286n6
- Mayo, Deborah, 79n1, 492n3, 536, 546n20
- Mayr, Ernst, 996n3, 1002n38
- McCarthy, Thomas A., 406n35
- McCarthy, Timothy, 729–31, 732, 733, 742n15, 744n16, 787, 788
- McClintock, Barbara, 248, 253n17, 980
- McGinn, Colin, 850
- McGuinness, B. F., 900n3
- McMullin, Ernan, 84, 137n17, n20, 138n28, n29, n35, 220, 224, 225, 228, 230–35, 236, 239, 250, 406n35, 478n8, 1133n8
- Meixner, John, 803n14
- Mellor, D. H., 544n5, 623n8, 806, 848, 848n, 849, 854, 857, 858, 859, 861, 880, 887, 888, 890–92, 893, 894–96, 898–99, 900n10, 1172n
- Mendel, Gregor, 40, 55, 58, 974–95, 976, 977, 1028, 1031, 1032, 1181, 1251–52, 1118, 1175. *See also* (in subject index) genetics, Mendelian
- Mendeleev, D. I., 919
- Messiah, Albert, 871–73
- Metzger, Hélène, 100n3, n8
- Meyerson, Émile, 101n9, 300n20
- Michalos, Alex, 967
- Michelson, A. A., 13n, 178, 184
- Milhaud, G., 271
- Mill, John Stuart, 160–61, 423–24, 462, 502, 503–5, 539, 545n11, n12, 727, 731, 733, 735, 738, 740, 742n15, 788, 822, 869, 870, 873, 874, 886, 901n12, 1107–8, 1166, 1296, 1303, 1304
- Miller, Arthur I., 24n, 1206n15
- Miller, David, 352n49
- Miller, Roger, 1168n12
- Millikan, Robert, 178, 1096, 1156, 1164
- Milton, John, 1071
- Minkowski, H. Hermann, 316
- Mitroff, Ian, 50
- Mittelstrass, Jürgen, 348n7
- Molière, 95n, 97
- Molnar, G., 844n9
- Moore, John N., 47n16
- Morgan, T. H., 971, 974, 975, 980, 1028, 1031
- Morgenbesser, Sidney, 605n3, 945n8, 968n10, n13, 969n21
- Morley, E. W., 13n, 178, 184
- Morris, Charles, 1107, 1109
- Morris, Henry M., 42, 46n12, 47n16, n20, 57, 61n10
- Moses, 45
- Motley, Michael T., 563, 582n24
- Motte, A., 82n20, 151n, 1112n12
- Muller, H. J., 971, 980, 985, 988
- Munitz, Milton, 968n14
- Murdoch, Dugald, 116n, 924n
- Musgrave, Alan, 12n, 85n2, 118n3, 167n1, 347n2, 411n3, 478n8, 482, 545n19, 546n29, 547n34, 807n1, 905n, 1050, 1051, 1112n11, 1113n15, 1220, 1241–46, 1272–77, 1279–80, 1284–85, 1288n31, 1289n35
- Naess, A., 925
- Nagel, Ernest, 45n5, 139, 403n13, 459n11, 546n28, 799, 807n2, 843n6, 903–4, 926, 930, 950, 953, 954–58, 959, 960, 961, 966, 967n2, n6, 968n7, n8, n14, n17, 972, 997n6, n7, 1005–22, 1023, 1024–28, 1039–42, 1044n1, n3, n6, n7, 1045n14, 1052, 1062n2, n3, 1172n, 1222n, 1285n1, 1292, 293, 1294, 1298, 1300, 1301, 1305

- Nash, L. K., 12n
 Nehamas, Alexander, 348n6
 Nelkowski, H., 1206n14
 Neumann, F. E., 261, 262, 359
 Neurath, Otto, 1301
 Newcomb, Simon, 304, 601, 606n18
 Newton, Isaac, 5, 10n4, 21–22, 24–25, 29, 34, 39–40, 50–51, 63, 68–71, 79n1, 80n11, 81n13, n14, 82n20, 91–96, 98, 99, 123–24, 150–51, 158, 161, 179, 186n, 225, 234, 251n7, 262, 263, 265, 266–70, 297, 303, 304, 306, 308, 314, 341, 357, 359, 363, 364, 365, 375, 382, 392, 398, 402n7, 406n34, 447, 451, 452n, 457, 461, 462, 472, 475, 479n19, 506, 517, 528, 531, 559, 565, 568, 574–75, 584, 586, 599, 655, 656, 687, 688, 771, 796, 805, 813n, 817, 847, 850, 881–82, 897, 898, 910, 929, 932, 939, 940, 941, 947n26, 952, 960, 968n14, 1007–8, 1014, 1016, 1038, 1046n15, n20, 1091–92, 1101, 1102–4, 1112n12, 1113n13, 1118, 1159, 1183, 1298, 1302, 1307, 1309. *See also* (in subject index) Dynamics; Gravitation/Gravity; Hypotheses, Newton's rejection of; Induction, Newton's method of; Law(s), of motion; Mechanics, classical; Methods/Methodology, Newtonian; Optics; Physics, Newtonian
 Newton-Smith, W. H., 321, 347n3, 545n17, 1114, 1115, 1117, 1125, 1129, 1131, 1133n8, n9, 1140, 1141–45, 1147, 1184n6, 205n5, 1208n30, 1253, 1256–59, 1281, 1287n19, 1288n22
 Nickles, Thomas, 92n, 411n2, 544n7, 545n14, 744n20, 904, 967n2, n3, 969n23, 1022–28, 1041–42, 1206n15
 Niiniluoto, Ilkka, 673n26, 1114, 1133n8, n10, 1206n11
 Noah, 49–50, 58, 1126
 Nola, R., 348n7
 Nye, Mary Jo, 1045n9
 Ockham, William of, 672n17
 Oedipus, 6, 9n3
 Oersted, Hans Christian, 560, 561, 561n
 Okruhlik, Kathleen, 85, 208n15, 214, 244–50
 O'Connor, D. J., 1288n31
 Ohm, G. S., 726, 844n7
 O'Leary-Hawthorne, John, 1243
 Oppenheim, P., 694n2, 802n1, 843n1
 Ortony, Andrew, 136n4
 Osiander, Andreas, 1188n
 Oster, G., 1002n39, n40
 Ostwald, Wilhelm, 455
 Overton, Judge William R., 45n3, n5, 47n23, 48–52, 54–58, 60, 60n1, 75–76, 81n18
 Pais, Abraham, 1222
 Pannekoek, Anton, 81n13
 Papineau, David, 677n1
 Pardies, Ignace Gaston, 941, 947n26
 Parker, Gary E., 45, 47n22, n24
 Parmenides, 8
 Partington, J. R., 100n3
 Pascal, Blaise, 310, 318n3, 448, 448n, 1205n7
 Pauli, Wolfgang, 747, 747n, 1195
 Peano, Giuseppe, 317
 Pears, D. F., 900n3
 Pearson, Karl, 455, 519, 807n1, 1166
 Peirce, Charles S., 151, 152, 292, 337, 1065, 1075, 1291
 Périer, Florin, 448, 448n
 Perrin, Jean, 492n6, 936, 947n29, 1013
 Perry, John, 865
 Peters, Donald, 187n8
 Peters, J. A., 996n1, n2, 1001n25
 Pickering, Andrew, 1163
 Pinch, Trevor, 79n1, 411n2, 546n29
 Pitt, Joseph C., 348n7, 546n20, 1204n1
 Planck, Max, 382, 431n6, 452, 920, 946n23, 951, 964, 966, 970n35, 1144, 1258
 Plato, 39, 1157, 1166n
 Podolsky, Boris, 116n
 Poincaré, Henri, 262, 271, 274, 275, 279n11, 303n, 306, 307, 309, 310–11, 316, 362, 364, 398, 813, 813n, 1188, 1294
 Poisson, Siméon Denis, 528, 529, 533, 565, 572, 577
 Polanyi, Michael, 321
 Pons, Stanley, 468–69, 469n, 472
 Popper, Karl R., 2, 11–15, 17, 17n2, 18n17, 22–26, 30, 33–34, 45n5,

- Popper, Karl R. (*cont.*)
 46n9, 52, 55, 57, 63–71, 78,
 81n14, 82n20, n21, 165, 186n3,
 188n25, 255, 332, 335, 337,
 352n49, 360, 363, 387, 402n6,
 406n31, 410, 434–43, 456–57,
 463, 477n3, 481, 486, 496, 502,
 505–10, 511, 519, 522–23, 539–
 40, 542, 545n13, n15, n17, 575,
 640, 652, 843n5, n6, 848, 848n,
 925, 926, 944n1, 945n6, 946n18,
 947n29, 1016, 1062n2, 1094,
 1111n4, 1113n13, 1118, 1133n9,
 1136, 1142, 1143, 1146, 1187,
 1124, 1285n1, 1288n23, 1294,
 1295, 1297, 1299, 1310
 Powers, Jonathan, 13n
 Prescott, C. Y., 1161, 1164, 1168n7,
 1178
 Presswood, S., 18n5
 Price, Derek de Solla, 352n48
 Priestley, Joseph, 97n, 102, 123, 143,
 145, 157
 Proust, Marcel, 430n
 Prout, William, 1118
 Ptolemy, 13n, 17, 28, 31–33, 68, 72,
 73, 84, 103–5, 108, 132–34, 220–
 21, 229–30, 234, 297, 340, 454,
 474–75, 1078
 Putnam, Hilary, 380, 381, 404n21,
 n23, 605n3, 806, 857, 858, 867,
 888, 892–96, 898–99, 900n8,
 1000n20, 1066, 1082–84, 1087n27,
 1099, 1101, 1114, 1115, 1117–20,
 1121–22, 1123, 1125, 1127, 1128,
 1129, 1131, 1132n1, n2, n3, n4,
 1133n7, n8, 1134n15, 1137, 1140–
 41, 1142, 1155–56, 1172–73,
 1174, 1180, 1187, 1196, 1204n1,
 1207n27, 1213, 1240, 1255, 1262,
 1265–66, 1279, 1283, 1288n9,
 1299, 1306
 Quine, W. V., 30, 156, 255–56,
 284n, 294n, 299n8, 300n11, nn.
 18–19, 302, 311–17, 318n4, n5,
 n6, 321, 324–35, 336, 338, 339,
 342, 343, 344, 345, 346–47,
 348n9, 349n10, n12, nn14–16,
 350n22, 351n23, nn27–30, n35,
 352nn, 36–39, 353n50, 365–76,
 378–79, 380, 381, 382–89, 397,
 398–400, 401, 402n9, n10, n11,
 403n14, n15, n16, 404n17, 405n25,
 n26, n27, 459n9, 647, 931, 1291,
 1294, 1296, 1298, 1300, 1309. *See*
also (in subject index) Duhem-
 Quine thesis; Quine thesis
 Quinn, Philip L., 81n18, 328
 Quinton, Anthony, 1220n
 Railton, Peter, 677, 747n, 785, 790–
 99, 801–2, 1295, 1303
 Ramsey, Frank P., 560, 623n8, 635,
 820, 861, 862, 890
 Randi, James, 188n14
 Rayleigh, Lord (J. W. Strutt), 492n6
 Redhead, Michael, 742n5
 Reese, W. L., 100n22
 Reichenbach, Hans, 427, 427n, 428,
 431n3, 455–56, 507, 519, 520–21,
 539, 542, 544n7, 551, 553, 580n2,
 640, 641, 672n12, 680, 718n5, 768,
 807n2, 1077, 1079, 1294, 1301
 Resnik, David B., 1051, 1263–67,
 1283
 Rescher, Nicholas, 137n21, 1002n37
 Richardson, Robert C., 188n32
 Richter, Burton, 1197
 Riemann, Bernard, 316
 Roentgen, Wilhelm, 87, 87n, 122
 Rogers, E., 459n6
 Roller, Duane H., 458n2, n3
 Rolfs-Hansen, Nils, 969n21
 Rorty, Richard, 255, 1204n1, 1207n27
 Rose, Steven, 188n32
 Rosen, Nathan, 116n
 Rosenberg, Alexander, 1185n16, 1251
 Rosenfeld, Léon, 948n31, 949n41
 Rosenkratz, Roger D., 406n31, 596–
 97, 606n17, 621n3, 651, 653–54,
 673n23, n24, n31
 Rowell, Thelma, 178
 Ruben, David-Hillel, 676–77, 725n,
 744n17, 778, 784–85, 786, 787,
 788–90, 800, 801, 802, 803n15,
 838n
 Rudner, Richard, 704n6
 Rumford, Count (Benjamin Thomp-
 son), 1121
 Ruse, Michael, 2, 45n2, n4, 46n6, n7,
 n8, 46n16, 61n10, 76, 77, 81n18,
 82n22, 996n3
 Russell, Bertrand, 281, 293, 294n,
 317, 378, 817, 1059n, 1184n4,
 1301, 1304
 Rutherford, Ernest, 919
 Rydberg, Johannes Robert, 965

Ryle, Gilbert, 734, 807n1, 820
Rynin, David, 925, 946n20

Sagan, Carl, 1

Salam, Abdus, 1160

Salmon, Merrilee, 677n1

Salmon, Wesley C., 242, 406n31,
410, 434, 444n2, 477, 508–10,
520, 528, 539, 540, 544n6, 545n18,
550, 561n, 562, 580n4, 581n10,
582n26, 605n4, 621n3, 634–35,
638–46, 668, 671n8, 672n12, n13,
n14, 677n1, 724–25, 726, 728,
734, 742n7, 765n5, n6, 784, 785,
786, 787, 792, 799, 800, 801, 802,
981n, 1047n24, 1077, 1079, 1149,
1150, 1151n15, 1184n14, 1288n23,
n25, 1296, 1303

Savage, Leonard J., 591, 621n3, 635,
648–50

Savage, C. Wade, 645

Schaffer, Simon, 546n29

Schaffer, Kenneth, 187n13, 950, 957,
958–59, 960–62, 968n13, 969n26,
996n3, 997n7, 1002n33

Scheele, Carl, 97n

Scheffler, Israel, 118n5, 251n5, 702,
717n2, 1286n2

Schelling, Friedrich, 561n, 678

Schiebinger, Londa, 197

Schilpp, Paul A., 4n, 546n28,
581n14

Schlick, Moritz, 820, 1285n1, 1301

Schrödinger, Erwin, 586, 755, 865,
1195, 1196, 1207n22, 1269

Schwartz, Stephen P., 900n7

Scriven, Michael, 697–99, 702,
703n2, n4, 704n6, n7, 726, 734,
735–37, 744n21, 775, 776–77,
799, 946n20, 1062n2

Seidenfeld, Teddy, 621n3, 622n7,
623n9, 624n16

Selby-Bigge, L. A., 745n24

Sellars, Wilfrid, 1062n2, 1063n7, n9,
1065, 1071, 1075, 1076, 1077,
1079–82, 1085n1, 1086n17, 1117–
18, 1132n5, 1133n8, 1184n2,
1204n1, 1237, 1238–40, 1279

Sextus Empiricus, 1073, 1129

Shapere, Dudley, 118n4, 125–26,
970n36, 991n12, 997n8, 1165,
1176

Shapley, Harlow, 445, 446n

Shils, E., 918n

Shimony, Abner, 444n4, 556, 561–
62, 593, 621n3, 641, 1114

Shugart, C. G., 1206n16

Simons, Gary, 187n8

Skinner, B. F., 367

Sklar, Lawrence, 950, 957, 958–59,
960, 967n5, 969n18, 997n8, 1026

Skyrms, Brian, 417, 544n6, 621n3

Slezak, Peter, 406n35

Slusher, H. S., 57n16

Smart, J. J. C., 865, 865n, 1075,
1077–79, 1078, 1084, 1133n8,
1137, 1177, 1237–38, 1279

Smith, Joel, 583n40

Smokler, H. E., 623n8

Smoluchowski, Marian von, 947n29,
1013

Snell, Willebrord, 313, 382, 446, 447,
448, 457, 511, 688, 770, 1297

Snyder, Laura, 411, 482, 525–33,
534, 535, 541, 542–43, 547n35,
669, 1298

Sober, Elliott, 80n2, 1002n38

Soldner, J. G., 606n20

Sommerfeld, Arnold, 969n19, 1195

Sosa, Ernest, 741n4

Spielman, Stephen, 622n4, 623n12,
624n14, n16

Speck, Josef, 968

Stachel, John, 1204n14

Stahlmann, William Joseph, 18n15,
34

Stalin, Joseph, 252n14

Stalker, Douglas, 405n28

Stanley, Steven M., 1002n38

Stauffner, R. C., 1044n1

Stebbing, L. Susan, 905n

Steel, Daniel, 674n35

Stein, Gertrude, 477

Stein, Howard, 1194

Stern, Otto, 1156

Stoney, G. Johnstone, 1156

Strauss, M., 970n37

Strawson, P. F., 403n15, 1083

Strickberger, M. W., 999n16,
1001n24, n25

Stuart, Walter W., 188n14

Stump, D. J., 406n35

Sturtevant, A. H., 980, 984

Summers, Montagne, 948n32

Suppe, Frederick, 967n3, 997n6,
1086n17, 1285n1

Suppes, Patrick, 459n11, 563,
582n25, 606n17, 621n3, 622n4

- Svedberg, T., 936
 Swain, Marshall, 1086n12
 Swartz, Norman, 807n2
 Swedenborg, Emanuel, 39
 Swift, Jonathan, 29
 Swinburne, Richard G., 544n6,
 673n31, 674n32
 Swoyer, Christopher, 807n2
 Sybenga, J., 998n14
- Tarski, Alfred, 459n11, 1212–15,
 1273–74, 1288n31
 Taylor, C. W., 136n1
 Teller, Paul, 606n17, 647, 671n7,
 1206n10, 1207n26
 ter Haar, D., 946n12
 Thagard, Paul R., 2, 32, 35n2, n5, 71,
 72–74, 78–79, 82n21, 1076,
 1086n11, 1295
 Thompson, Silvanus P., 100n1
 Thomson, J. J., 490, 491, 492n6,
 537–38, 870, 1156, 1161
 Thorndike, Lynn, 18n13, n15, n14,
 n15, 19n18, 34
 Tichy, Pavel, 352n49
 Ting, Samuel, 1197
 Todhunter, I., 353n51
 Tooley, Michael, 807n2, 858, 859–
 60, 861, 862, 885, 1310
 Toonder, J. C., 35n3
 Torricelli, Evangelista, 448, 448n
 Toulmin, Stephen, 140, 697–98,
 717n2, 807n1
 Traweek, Sharon, 189n35
 Tuana, Nancy, 207n1
 Turnbull, Robert G., 403n13
 Tweedale, Martin, 900n4
- Uehling, Theodore E., Jr., 544n4
 Uhlenbeck, George, 1156
 Urbach, Peter, 406n32, 478n15, n17,
 611, 621n3, 657, 673n28, 807n2
- Van Cleve, James, 544n4, n5, 1160n
 van der Waals, Johannes Diderik, 652
 van der Waerden, B. L., 1207n19
 van Fraassen, Bas C., 132, 255,
 411n3, 621n3, 807n1, 865, 887,
 889, 1050, 1085n1, 1086n22,
 1088–1111, 1111n1, n2, n4,
 1112n7, 1113n4, 1138–39, 1145–
 46, 1167, 1179, 1188, 1196,
 1207n27, 1209–10, 1232–46,
 1254–55, 1259–60, 1267, 1272,
 1278–81, 1286n10, 1291, 1293,
 1296, 1303, 1309
 Vavilov, Nicolai Ivanovich, 26
 Velikovsky, Immanuel, 1, 734, 530,
 562
 Vigier, J. P., 928
 Volta, Alessandro, 560
 Voltaire, 28
 von Neumann, John, 380
 Vuillemin, J., 313, 318n4
- Waismann, Friedrich, 1301
 Wallis, John, 96
 Wartofsky, M. W., 404n21
 Waters, C. Kenneth, 1033
 Watkins, John W. N., 33, 434, 437–
 39, 441–42, 509–10, 545n18, 925,
 1112n5
 Watson, James D., 187n7, 903, 971,
 975, 985, 988, 996, 996n2, 998n9,
 999n16, 1001n31, 1028, 1035
 Watson, John, 367
 Wayne, Andrew, 674n35
 Weber, Max, 252n13
 Wegener, Alfred, 29, 1119
 Weinberg, Steven, 187n7, 583n39,
 1160, 1207n18
 Weinberg, Wilhelm, 39
 Weinert, Friedel, 807n1
 Weingartner, Paul, 970n37
 Wessels, Linda, 1002n39, 1047n25,
 1207n20
 Wessels, N. K., 1002n36
 West, J. A., 35n3
 West, Mae, 949n38
 Westfall, Richard S., 448n
 Westman, Robert, 137n17
 Wettstein, Howard K., 544n4
 Wheller, John A., 1
 Whewell, William, 130, 150, 151,
 337, 462–63, 467, 481, 486, 1100
 Whitcomb, John, Jr., 42
 White, Morton, 300n22
 Whitehead, Alfred North, 378
 Whitehouse, H. L. K., 1001n25
 Whittaker, E. T., 101n10, 547n31,
 n36, 946n22
 Wickramasinghe, N. C., 44
 Wiedemann, G. H., 957, 958, 959,
 1026
 Wien, Wilhelm, 1205nn7
 Wiener, Otto, 261, 262, 359, 360
 Wiener, Philip P., 100n2, 431n5,
 802n3

- Wigner, Eugene, 13n
 Williams, Bernard, 415
 Williams, D. C., 718n3
 Williams, L. Pearce, 188n19
 Wilson, C. T. R., 739
 Wirth, L., 918n
 Wimsatt, William, 996n4, 1000n20
 Wittgenstein, Ludwig, 15n, 231,
 294n, 321, 369, 659, 807n1, 890,
 900n3, 1059n, 1063n9, 1223n2
 Woodward, James, 623n10, 624n15,
 726
 Woolf, Harry, 1205n7
 Worrall, John, 347n3, 477, 513, 514,
 527, 528, 529, 547n36, 582n33,
 1111n2, 1113n14
 Wren, Christopher, 96
 Wu, Chien-Shiung, 13n, 449n
 Yang, Chen Ning, 13, 449
 Young, Thomas, 40, 265, 365,
 401n4
 Zabłudowski, Andrzej, 406n29
 Zahar, Elie, 482, 513, 546n29,
 1111n4
 Zecha, G., 970n37
 Zenker, Wilhelm, 262

SUBJECT INDEX

Note: Headings and subheadings marked with an asterisk (*) correspond to entries in the Glossary.

- *Abduction. *See* Inference, to best explanation
- Accommodationism. *See* Confirmation, and accommodationism
- Accuracy, predictive, 437, 463, 1078, 1142. *See also* Prediction(s)
 - as criterion of theory assessment, 103–4, 110–12, 114, 116, 117, 120, 125, 127, 129–30, 132–35, 157, 160, 182, 168n12, 212, 226, 228, 232–34, 247, 251n8, 339, 391, 401, 552, 575, 645, 1066, 1118, 1125, 1128–29, 1137, 1147, 1166, 1233
 - and explanatory power, 104, 133
- Action at a distance, 20, 77, 95n, 96–97, 225, 796
- Ad hoc devices. *See* Hypotheses, ad hoc
- Aether. *See* Ether theories
- Alchemy, 28, 29, 72
- *Analytic statements, 80n7, 280–83, 285–93, 295–96, 311–13, 366–73, 399, 404n17, n22, n23, 427, 812–14, 827, 847, 850, 1010, 1039
- Analyticity. *See* Analytic statements; Analytic-synthetic distinction
- Analytic-synthetic distinction, 280–96, 311–12, 366–74, 398–99, 404n17, n18, 647
 - and conventionalism, 812–14
 - and definition, 283–85, 289, 311, 369, 372, 404n22
 - Frege's doctrine of, 368–69
 - Kant's doctrine of, 368–69
 - and logical truths, 282–83, 404n22
 - and meanings, 281, 312, 366, 369–71
 - and reductionism, 295
 - and semantical rules, 289–92, 373
 - and state-descriptions, 282–83, 372, 403n14
 - and synonymy, 282–89, 292–93, 372, 403n14
 - and verification theory of meaning, 292–95
- Ancestry, human, 41, 44, 50, 75
- Anomaly/anomalies, 23, 24, 31–32, 40, 64, 71, 90, 93, 120, 123, 136n1, 215, 218, 251n6, 345, 571
 - of perihelion of Mercury. *See* Mercury, anomalous perihelion of
- Anthropology, 194–95, 246, 250
- *Antirealism, 867, 1050–51, 1062n5, 1067–70, 1078, 1081–83, 1088–90, 1094, 1100, 1110, 1129–30, 1139, 1145–48, 1153, 1166–67, 1176, 1181, 1186, 1188, 1195–1200, 1202–3, 1204n1, 1207n25, n27, 1209–12, 1221, 1223, 1227–30, 1232–1238, 1241, 1243–45, 1252, 1259–63, 1267–72, 1276, 1289n35. *See also* Idealism; Instrumentalism
- and acceptance, 1067, 1074–75, 1077
- and fundamental laws, 867, 896–97
- influence of Kuhn on, 132
- and success of science, 1084, 1099–1100, 1139, 1147, 1245, 1253–55
- and truth, 1067, 1137–39, 1214–15, 1223n1, 1287n17
- Applied sciences. *See* Science(s), experimental
- *A priori knowledge. *See* Knowledge, a priori
- Aristotelianism/Aristotelians, 10n4, 95, 123–24, 127, 146, 237, 281, 311, 448n, 593, 1103, 1219, 1224n3. *See also* Physics, Aristotelian/scholastic
- Arithmetic, 377, 381, 1267–68
- Astrology, 1, 4, 5, 8, 10n4, 15–17, 27–35, 63, 67–68, 72–74, 439, 463
- Astronomy, 14, 86–87, 114, 273–74, 304, 683 *See also* Revolution, Copernican

Astronomy (*cont.*)

and astrology, 5, 16–17, 63, 1144, 1259

Brahe's, 13n, 222–23, 270, 463, 464, 568

Copernican (heliocentric), 13n, 17, 20, 26, 68, 84, 94, 103–6, 108–9, 111–12, 132–35, 158, 177–78, 220–23, 226, 228–30, 232–34, 340, 454, 567–68, 570–71, 586, 599, 645, 944n2, 1077–78, 1224n3, 1237–38

and predictions, 16, 111–12, 129, 133–34, 232, 462

Ptolemaic (geocentric), 13n, 16–17, 28, 84, 94, 103–5, 108–9, 111, 132–35, 158, 220–22, 226, 229–30, 232–34, 340, 454, 567–68, 586, 1071, 1077–78, 1100, 1126, 1224n3, 1235, 1237–38, 1245

Atonism, 65, 97n, 153, 154, 238, 310, 376, 584, 767, 803n5, 927, 1055, 1118–19, 1121, 1126, 1141, 1166–67, 1227, 1229, 1255

logical, 294, 1059

Australopithecines, 44

Authority, intellectual, 181, 183–85, 214, 244

*Axiology, *See* Paradigm(s), Values; components of

*Bayesian confirmation theory, 80n11, 242, 332, 390, 510, 523–24, 530, 549–50, 553–80, 587–605, 605n5, 606–11, 627–71

See also Theory/Theories, choice, Bayesian algorithms for

and coherence condition, 471, 557, 560, 577–78, 538n42, 589, 601–3, 609, 611, 613, 615, 617, 619, 621, 623n8, 635–38, 641, 646, 647–48, 657, 658, 664, 667, 669

and conditionalization, 578, 588, 597, 600–602, 636–38, 646, 647–49, 657, 668

and convergence theorem, 590–91, 637, 642, 648–50, 668

criticized, 584–605, 646–59

and diversity of evidence, 512, 611–13

and expectedness, 550, 556, 564–66, 570, 577, 609, 633–34, 642–44, 661, 667–68

and *likelihoods, 554, 558, 564–74, 577–78, 594, 597, 599, 633–34, 638, 642–46, 648–49, 651, 653–54, 661, 667–68

and old evidence, 528, 550, 599–604, 623n10, 634, 643, 654, 656–59, 669

and *posterior probabilities, 554, 556, 566, 573, 577–78, 597, 582n29, 627, 632–33, 639, 642, 644–45, 647, 661, 667–68, 672n13

and *prior probabilities, 554–56, 559–65, 570, 574, 576, 578–79, 582n29, 586–87, 591–95, 597, 602, 609, 620, 627, 633–37, 639–48, 651–52, 657, 661, 663, 667–68, 670, 672n13

and raven paradox, 608–11, 614, 615–16, 659–63, 670

and scientific realism, 614, 655

subjective, 471–72, 474, 475, 479n21, 531, 542, 587, 635–36, 641, 667–68

therapeutic, 550, 607–21, 659–67, 670–71

*Bayes's theorem, 70, 109, 464–65, 472, 528, 549–50, 551, 553–59, 564–66, 568–70, 577–79, 581n16, 609, 628–34, 638–40, 642, 647–48, 661, 667–68, 797

derived, 629–31

Belief, 20, 69, 386, 765n8

and acceptance, 1066–67, 1069, 1074, 1089, 1093, 1232–34, 1241

degrees of, 471–72, 474, 558, 560, 577–78, 585, 587–92, 598–604, 605n5, 607–11, 613–16, 618–19, 621, 622n4, 628, 634, 635–36, 638, 640, 647, 649, 654–55, 657–60, 665–68, 670, 1067, 1200, 1235

explanation of, 393–95, 917n, 1265

reasons for, 467–69

in theories, 1096–97, 1235, 1241

web of, 296–99, 327–29, 335, 373–74

- Bias, 116, 175–76, 205, 213, 252n14
 in biology, 192–199, 206–7, 214, 244, 245–46, 249
 elimination of, 85, 205, 245
 gender, 85, 178, 183, 192–99, 201–7, 213–14, 244, 245–46, 248, 249
 in physics, 206
 racist, 85, 183, 213, 244
- Biology, 33, 43, 44, 114, 151, 154, 175, 187n5, 192–199, 203, 237, 245–46, 250, 251n1, 252n14, 677n1, 679–82, 776, 797, 801, 813, 865–68, 904, 906, 911, 991, 992–96, 1028, 1042–43, 1049, 1074, 1095, 1097, 1236, 1244
 molecular, 903–4, 971–73, 975–80, 985–86, 991–96, 1002n35, 1028, 1030, 1035, 1037, 1042.
See also Genetics, molecular
- Brownian motion, 164, 934–36, 1013
- Calcination, 12, 771–72
- Cathode rays, 87, 123, 490–91, 537–38, 1156
- Causation/Causes, 423–24, 503–5, 738, 741n4, 797, 811, 814, 871, 1106, 1170, 1174, 1176–78, 1220, 1220n, 1262, 1282–83. *See also* Explanation, and causation; Necessity, causal
 composition of, 869–76
 constant-conjunction theory of, 738, 806, 879, 1171
 and explanation. *See* Explanation(s), and causation
- Hume's theory of, 879–80, 898, 1171, 1176, 1264, 1288–26
 and the strong programme in sociology of science, 343–44, 393
- Certainty
 and experimental method, 264–66, 305, 363, 402n5
 as goal in science, 22, 152–53, 237
- Chance. *See* Determinism/Indeterminism
- Chemistry, 16, 29, 33, 39, 44, 51, 72, 89, 97, 114, 123, 142–43, 151, 165–66, 237, 238, 257, 259–60, 271, 275, 313, 394, 401n2, 865, 874, 906, 911, 919, 976, 988, 990–91, 1002n34, 1005, 1007, 1028, 1049, 1056–57, 1118, 1121, 1126, 1128, 1180–81, 1182, 1266. *See also* Oxygen theory of combustion; Phlogiston theory of combustion; Revolution, chemical
 explanation of qualities in, 92, 98, 99, 104, 114–15, 165–66
- Cognitive science, 179
- Coherence, 129–31, 134–35, 345, 618, 917
- Cold fusion, 468–69, 472
- Comets, 24–25, 66
- Communication, 124
 between rival theories, 117, 135, 136n5, 173
- Community, scientific, 41, 84, 88, 95, 98–99, 102, 110, 112–13, 141, 144, 151, 158, 173, 174–76, 179–80, 181–86, 205–6, 214, 245, 248–50, 251n2, 310, 338–39, 552, 829, 831, 1097, 1173, 1192
- Conditionals, subjunctive/°counterfactual, 821–22, 830, 832–33, 838, 841, 851–52, 854–55, 859, 873, 882–83, 885, 976, 1032, 1070
- Confirmation, 31, 40, 43–44, 73–74, 78, 79, 83, 135, 203–4, 229, 238, 241–42, 247, 257, 292–93, 295, 329, 351n25, 361, 367, 383, 389–90, 409, 421, 445–453, 488, 502, 504, 510–38, 540–43, 549–50, 556, 577, 584–85, 587, 619–20, 623n9, 635, 666–67, 871, 923, 925, 927, 938, 941–42, 943, 953, 1096, 1118, 1200, 1203, 1231, 1285n1. *See also* Evidence; Induction; Prediction(s); Testing/Test(s); Theory/Theories, empirical support for; Theory/Theories, theoretical support for
 absolute criterion of, 546n23
 and accommodationism, 514–15, 525–26
 a priori theories of, 482, 534, 541, 557
 and background beliefs, 180, 195, 203–4, 214, 240–43, 424, 449, 479n21, 554, 556, 580, 582n27, 609, 622n4, 633, 656–57, 673n31, 700. *See also*

Confirmation (*cont.*)

- Evidence, and background information; Hypotheses, auxiliary
- Bayesian. *See* Bayesian confirmation theory
- bootstrap, 204, 469–70, 482, 588, 646–47, 655, 661
- and converse consequence condition, 517, 541, 844n10
- degrees of, 390, 458, 524, 550, 557, 608–11, 616, 627, 639, 645, 654, 661, 665, 666, 670, 689–90, 708, 711, 782
- and entailment of evidence, 515, 553, 594, 638, 653, 658, 669
- and equivalence condition, 546n22
- and explanation. *See* Explanation(s), and confirmation
- and formalism, 514–16
- functions, 608, 617, 619, 666, 670
- and historical circumstances, 547n34, 602–3
- hypothetico-deductive, 204, 422–23, 482, 518, 541, 549, 559, 579, 593, 598–99, 603, 627, 638–40, 655, 659, 661, 668. *See also* Induction, hypothetico-deductive model of; Hypothetico-deductive model
- induction, 384, 389, 400, 420n5, 409–10, 469, 505, 509, 510, 516–17, 531, 540, 638, 640, 645, 708, 1122
- instantial model of, 502–3
- by “new” evidence, 166, 239, 332, 402n5, 410, 417–18, 450–51, 460–63, 512, 525, 527–31, 535–36, 538, 621. *See also* Prediction(s), novel; Predictionism/Predictivism
- objectivity of, 613, 616, 665
- and observation, 29, 220, 1060
- paradoxes of, 550, 608, 618, 660, 666
- by positive instances, 5–7, 8, 63, 337, 388, 389–90, 406n31, 420–22, 435, 502–3, 518, 542, 591, 660, 673n31, 821, 827, 838. *See also* Induction, enumerative; Induction, straight rule of
- positivist account of, 172, 240, 367
- and prediction, 212, 220
- and probability, 506, 541
- and raven paradox, 421–22, 503, 527, 546n22, 550, 608–11, 614, 615–16, 659–63, 670
- *relevance criterion of, 464–65, 471, 478n10, 527–28, 597, 619, 627–28, 654, 656, 658, 660, 665, 667, 834, 844n10
- *satisfaction criterion of, 252n12, 482, 517, 627
- and special consequence condition, 516–18, 541, 546n23, 844n10
- of statistical theories, 503, 536, 553, 611, 621, 638, 668, 672n13
- of synthetic statements, 296
- and tacking paradox, 64, 422–23, 503, 544n10, 587, 597, 639, 1092
- of universal generalizations, 833–37
- and variety of evidence, 511–12
- Consensus, 240
 - formation of, 144, 146, 147–48, 150, 155–56, 158, 244, 496–97
 - objectivity as, 109–110, 227
 - truth as, 20, 69, 245
- Consistency, 636, 1021
 - as criterion for theory assessment, 103, 105, 111, 116, 129, 158, 168n12, 177–78, 182, 228–29, 238, 247, 338–40, 362, 391, 406n33, 552, 562, 576, 577–78, 583–82, 618, 640, 642, 645–46, 668, 917, 1109
- Contextualism, 213
- *Conventionalism, 7, 8, 306–9, 311, 312, 362, 364, 398, 508, 540, 808, 813, 1067, 1198–99, 1210, 1270
- Conventions, 306–9, 312–13, 364, 370–71
- Convergent realism. *See* Realism, convergent
- Conversion, 23, 117, 119, 120–21, 139, 141, 155, 218–19, 227
- Corpuscles. *See* Mechanico-corpuscularianism
- Correlations, statistical, 74
- *Corroboration, 7, 33, 390, 431, 435–43, 463, 505, 508–10, 540, 1133n10
- Cosmology, 187n7, 341–42
- Craniometry, 196

- Creation, divine, 41–42, 50, 57–58, 75
- Creationism. *See* Creation-science
- Creation-science, 1, 38, 41–45, 48–52, 54–55, 57–60, 65–66, 75–77
- trial (*McLean vs. Arkansas*), 2, 38, 46n10, 48, 54–58, 60, 80n6, 75–76, 184
- Criteria of theory choice, 14, 103–18, 120–35, 155, 164, 171, 173, 186, 187n13, 212, 224–26, 243–44, 248, 338–40, 342, 345, 386, 392, 395, 399, 410, 445–58, 510–24, 540–42, 552–53, 574–77, 640, 645, 1109–10, 1139, 1246. *See also* Methods/Methodology; Values
- as algorithms/rules, 107, 109–12, 135, 157, 224, 238. *See also* Theory/Theories, choice of, Bayesian algorithms for
- ambiguity of, 107, 111, 156–60, 220, 225–26, 339
- changes in, 114–15, 121–29, 220, 225, 231, 238
- disagreement about, 162, 225–26, 235–36
- extraempirical, 135, 343, 395
- imprecise and conflicting, 103–5, 110–12, 157–59, 160–61, 182, 212, 220, 226, 238, 391, 552–53
- incomplete, 106–7, 111, 113, 120, 153–54, 156, 157–58, 249, 391
- objective (shared) vs. subjective (individual), 105–10, 112–13, 116, 121, 157–59, 163, 168n12, 214, 225–30, 231–32, 250, 339
- and truth, 212, 232–34
- and underdetermination. *See* Underdetermination, of theories by methods/evidence
- as values, 110–15, 117, 120, 124, 125–35, 141–56, 157, 160, 162, 163, 167n2, 169n38, 173, 212, 224, 227–233, 235. *See also* Values, cognitive
- Criticism, 26, 175–86, 214, 243, 925, 936, 942
- of background assumptions, 85, 178, 184, 186
- and objectivity. *See* Objectivity, and criticism
- response of community to, 181, 182–83, 214, 244
- standards of, 181, 182, 214, 243–44
- transformative, 178, 181, 243
- Crystallography, 276
- Darwinism, 106, 118n8, 153, 154, 194, 225, 251n1, 697–98, 775, 1084, 1099–1100, 1138–39, 1147, 1219, 1240, 1245, 1254, 1279–82
- Darwin's finches, 43
- *Deduction, 416, 422, 427, 430, 495–96, 507, 540, 543n2, 564, 576, 584–85, 638, 686, 689–90, 695, 697, 707–8, 711, 717n2, 735, 756, 767, 770, 779, 829, 836–37, 874, 884, 906, 909, 976, 1007. *See also* Explanation(s), deductive-nomological; Explanation(s), deductive-statistical, Reduction(s), and logical derivability
- justification of, 414, 500–501
- Deductivism, 21–22, 69, 77, 82n20, 151n, 323–24, 363, 401n1, 428, 431n5, 435, 443, 732–33, 879, 1146. *See also* Falsificationism
- Definition(s), 271, 369, 891, 1228
- and analytic truths. *See* Analytic-synthetic distinction
- *bridge laws as, 913, 1010, 1022
- as explications, 283–84
- in formalized systems, 284–85
- hypotheses as, 271–72, 306, 364, 398. *See also* Conventionalism
- stipulative, 284
- and synonymy. *See* Synonymy, and definition
- *Demarcation of science from pseudoscience, 1–82
- criteria of, 2, 4, 7–9, 14–17, 17n, 22–26, 27, 29, 31–35, 39–41, 46n9, 48, 52, 54, 55–57, 63–79, 81n18, 128, 215, 505
- problem of, 9, 20, 23, 26, 34–35, 55, 71
- social/historical dimension of, 2, 27, 32, 33–34, 72, 78
- Demonstration (proof), 21, 69, 102, 125, 133–34, 772

Demonstration (proof) (*cont.*)

- direct, 266
- vs. probability, 22, 69–70, 152–53
- by reduction to absurdity, 261, 264
- *Determinism/Indeterminism, 741n4, 749, 761–62, 764n1, 927, 1081, 1105–6
 - biological, 192
 - and probabilistic explanation, 746, 752–53, 756–57, 761, 791–94, 797–98, 801–2
- Disciplinary matrix. *See* Paradigm(s), as disciplinary matrices
- Disconfirmation, 292–93, 295. *See also* Confirmation; Falsification
- Discovery, 90, 258, 943, 1065
 - *context of, 107–9, 113, 172, 193–94, 200, 203, 207, 240, 248, 250, 410, 429–30, 502, 504, 538, 544n7, 579, 638, 1053. *See also* Theory/Theories, generation of
 - logic of. *See* Logic, of discovery
 - rational reconstruction of, 429
- *Dispositions, 821–22, 854, 861, 897, 899
- Duhem problem, 639–40, 668
- Duhem thesis, 255, 302–4, 311, 313–14, 317, 374, 381. *See also* Duhem-Quine thesis; Holism; Quine thesis; Underdetermination
- *Duhem-Quine thesis, 195, 197, 246, 255–56, 260–64, 270, 273, 277, 295–97, 302, 314, 317, 327, 330, 342, 344, 350n19, 376, 379, 381–82, 397, 399. *See also* Duhem-Quine thesis; Duhem thesis; Underdetermination
- Dutch book arguments, 560, 590–91, 615, 635, 637–38, 647–48, 665, 667–68
- Dynamics, 89, 90, 91–94, 96, 269–70, 568
- Earth, the, 268, 567
 - age of, 75, 341
 - rotation of, 108, 474
 - shape of, 165, 341–42, 353n51, 392–93, 401
- Eclipse, observations of 1919, 4, 6, 12, 25, 63, 466, 477n7, 529, 600

- Electricity, 89, 97, 114, 123, 178, 561n, 680, 868–71, 896, 899, 1121, 1126
- Electrodynamics, 229, 310–11, 363, 577, 915, 1011, 1158–59, 1190
- Electromagnetism, 1159, 1165, 1168–10
 - theory of, 98–99, 123, 382, 547n38, 847, 906, 911, 915, 1019–20, 1092, 1167n10. *See also* Optics, and electromagnetic theory
- Electrons, 1153–65, 1167, 1175, 1177–78, 1181–82, 1201, 1214–16, 1230, 1262–63, 1265–66, 1270, 1275, 1282–83
- Embryology, 768, 994, 996, 1120, 1182
- *Empirical adequacy, 129–30, 182, 236, 247, 338, 618, 1050, 1067, 1069, 1075–78, 1089–90, 1092, 1095, 1105–6, 1109, 1147, 1233, 1235–38, 1240, 1241–42, 1244, 1253–54, 1257, 1259–60, 1278–79, 1281, 1286n5. *See also* accuracy, predictive and truth, 1088, 1094, 1096, 1109, 1112n7, 1207n27, 1210, 1241, 1272
- Empirical content. *See* Empirical significance
- Empirical method, 4, 63, 933, 936. *See also* Empirical world and objectivity, 180, 220
- Empirical significance, 296–97, 371, 375, 522–24, 542, 763, 770, 780, 799, 813–14, 903, 934, 936–37, 1019
 - unit of, 296, 314, 374–76, 399
- Empirical world, 39, 76
- *Empiricism, 139, 151, 173, 177, 292, 294, 296, 312, 366–68, 830, 871, 887–88, 897, 922–23, 925–28, 931–32, 936, 941, 943–44, 944n2, 1011–12, 1014, 1018, 1020–21, 1040, 1049–50, 1056, 1060, 1106, 1139, 1198–99, 1223n2, 1227–28, 1229–30, 1242, 1243, 1267, 1270, 1277–78, 1284

- *constructive, 1050, 1064, 1069, 1083, 1089–90, 1093, 1098, 1106, 1110–11, 1166, 1175, 1188, 1198–99, 1207–27, 1232–36, 1241–44, 1267, 1272, 1278–80
- dogmas of, 280, 292, 295, 312, 366–67, 399, 402n11
- feminist, 199, 205, 246–47, 249, 250–51
- *logical, 119, 312, 370, 427n, 550, 551–52, 577, 579–80, 640, 668, 923, 926, 1011, 1039, 1052n, 1227–29, 1232–34, 1254, 1277–78, 1285n1
- nondogmatic, 925, 1011
- strict, 1090, 1093
- Energy, conservation of, 89, 90–91, 106, 118n8, 682, 747n, 1165
- Entelechy, 676, 679–82, 768–69
- Epistemic values. *See* Values, cognitive
- *Epistemology/Epistemologies, 164, 167, 171, 320–21, 351n27, 611, 618, 659, 666, 903, 1067, 1101, 1114–17, 1125, 1132, 1136, 1166, 1170, 1173, 1180, 1203
- normative, 321, 325, 327
- standpoint, 199–200, 206, 246–47, 249, 250–51
- Equivalence, principle of, 449n
- Essences/Essential properties, 95, 281, 856–58, 892–96, 898
- Essentialism, 857–58, 893, 895, 898, 1102–5, 1113n13, 1212
- Ether theories, 13n, 82n20, 98, 178, 359, 490, 571–72, 645, 767, 803n5, 1118, 1120–21, 1127–28, 1141–42, 1190, 1248, 1255
- *Evidence, 83–84, 108, 164, 171, 177, 185, 238, 247, 255, 345, 445, 460–76, 477n2, 481–91, 525–38, 541–43, 585–86, 665, 938, 1122, 1172. *See also* Bayesian confirmation theory; Bayes's Theorem; Confirmation; Testing/Test(s)
- and background information, 178–80, 186, 194, 195, 203, 214, 240–43, 250. *See also* Confirmation, and background beliefs; Hypotheses, auxiliary
- changes in standards of, 123–24
- and confirmation, 44, 85n1, 627
- *contextualist analysis of, 85, 173, 179–80, 214, 240–42, 246, 250, 547n35
- empirical completeness of, 485–88, 535
- entailed by rival theories, 323, 384–85, 390, 423, 553, 594, 638, 1049
- “expert-relative” concept of, 473–75, 531–33, 542
- expectedness of. *See* Bayesian confirmation theory, and expectedness
- and explanation. *See* Explanation, and evidence
- and falsification, 56
- *historical thesis of, 410, 418, 451, 461, 466–67, 470–71, 475, 481–82, 484–86, 487, 525, 531–35, 542–43
- impersonal conception of, 467–75, 531–33, 542–43, 547n35
- objective vs. nonobjective concepts of, 411, 469–75, 482, 531–33, 542–43
- as objects and as statements, 241
- old, *problem of. *See* Bayesian confirmation theory, and old evidence
- positive-relevance view of. *See* Confirmation, relevance criterion of
- precision of, 410, 449, 512
- and prediction. *See* Prediction, and evidence
- and probability, 22, 69–70, 541, 707
- quantity of, 410, 445–46, 511, 540, 611, 621
- rejection of, 41, 326–27
- relations of theories to, 85n1, 214, 220, 323–25, 328–35, 383–89
- relevance of, to hypotheses, 178–79, 203, 550, 587, 593, 597–604, 646–47, 650, 654–55, 669
- reliable, 530–31
- and selection procedures, 483–86, 488, 490–91, 534–38, 543
- subjective Bayesian concept of. *See* Bayesian confirmation theory, subjective

Evidence (*cont.*)

- total, requirement of, 618, 711–13, 718n3, n7, 782–83, 1076
- variety of, 339, 410, 417, 446–449, 458, 483, 511–12, 524, 540, 549, 593, 608, 611, 612–13, 621, 622n7, 646, 650, 659, 663–64, 670

Evolution/Evolutionary theory, 8,

- 18n17, 38, 41, 43–45, 51, 52, 55, 58, 65–66, 75–76, 154, 184, 194–95, 430n, 680, 697–98, 775–76, 797, 1149, 1260, 1288n25 *See also* Darwinism

Exemplars. *See* Paradigm(s), as exemplars

Experience, 63, 128, 176, 923, 1011, 1018, 1049–51, 1227

- germaneness of statements to, 297
- immediate, 293–95, 367, 820, 1227
- and induction, 129–30
- and knowledge, 296

Experiment(s), 4, 11, 21, 43, 69, 96,

- 100, 159, 171, 174–75, 177, 211, 220, 257–66, 270–79, 302, 307, 355–62, 364–65, 376–77, 383–84, 397, 411, 424, 543, 829, 917, 923, 981, 1089, 1153–55, 1156–58, 1160–65, 1167n3, 1168n10, n13, 1169, 1171, 1176–79, 1180, 1261–67, 1282–83

on beta rays, 24n

on calcination, 12, 97n

on cathode rays, 87, 123, 490–91, 537–38, 1156, 1160

on cold fusion, 469

on collisions, 96

*crucial, 22–23, 26, 71, 108, 166, 220, 264–66, 278, 304–6, 310, 313, 359–61, 392, 397–98, 925, 936

on deflection of magnetic needle, 560–61

double slit, 40, 380

error in, 272–74, 537–38, 1162

explanation of, 1178, 1182

as intervention, 1156–57, 1169, 1174–75, 1183, 1261–64

methodological rules for, 257–59, 355

Michelson-Morley, 13n, 178, 184

on parity conservation, 13, 67

with PEGGY II, 1158–65,

1168n10, 1263, 1265, 1267, 1282

on polarized light, 261–62, 265, 359

Puy-de-Dôme, 448–49

on rhesus monkeys, 204

theory-dependence of, 259–64, 356–58, 376–77, 913, 1169, 1177, 1183, 1264, 1266, 1283

of testing, 260–63

torsion balance, 565

on velocity of light. *See* Light, velocity of

Experimental apparatus, 122–23, 259–60, 272, 356–58, 376–77, 1056, 1095, 1157–64, 1171, 1177–78, 1263, 1265, 1267, 1277–78, 1282

Experimental realism, *see* Realism, experimental

Experimental sciences. *See* Science(s), experimental

Explanation(s), 31, 65, 79, 83, 124, 129–30, 185, 225, 329–30, 386, 461, 526, 598, 675–802, 873, 875, 905n, 976, 990, 993–94, 1033, 1038, 1069, 1078–79, 1088, 1099–1111, 1148–51, 1153, 1169, 1199, 1227, 1263, 1287n12. *See also* Inference, to best explanation

adequacy conditions for, 769–72, 780, 794–96

as aim of science, 40, 236, 436, 439, 443, 675, 768, 770, 1238

as arguments, 77, 675–77, 685–86, 695, 732–36, 756, 767, 770, 779, 785, 788–89, 791, 794–95, 799–801, 911, 957, 1006, 1108, 1148–49

and causation, 133–34, 225, 234, 677, 687–88, 697, 722–23, 728–32, 739, 741n4, 742n15, 746–48, 756, 761, 770–71, 777, 785, 787–88, 791, 796, 800–802, 869, 876, 1106, 1179

and confirmation, 513–15, 526, 528–531, 541, 835–38, 842, 1149. *See also* Explanation(s), and evidence; Explanationism

*covering-law models of, 675–77, 686, 691, 785, 790, 799, 874, 978, 1148–49, 1285n1. *See*

- also* Explanation(s), deductive-nomological; Explanation(s), laws in
- *deductive-nomological, 77, 676, 686–94, 695–96, 700, 702–3, 706, 709, 715–16, 720–38, 746–48, 751, 754, 756–57, 767–74, 779–80, 785, 789–91, 793, 796, 799–800, 803n10, 903, 907–8, 911, 913, 926, 950, 953, 957, 976, 1006, 1008–9, 1039, 1044n6, 1148–50, 1260, 1282. *See also* Explanation(s), covering-law models of
- deductive-statistical, 706, 715–16, 750–51
- and description, 698, 806–7, 1104, 1108
- *D-N-P model of, 677, 752–64, 790–91, 794, 796–99, 801
- elliptical, 691–92, 772
- *epistemic conception of, 799–800, 802
- epistemic relativity of, 715–16, 751, 758–60, 780, 784, 793, 800–802
- and evidence, 329, 411, 460–67, 477n3, 479n23, 481–82, 486–90, 525–31, 534–35, 543, 1109. *See also* Explanation(s), and confirmation; Explanationism
- and generality, 737–38
- and high probability requirement, 690, 696–97, 704n6, 706, 750–52, 762–63, 775, 779–81, 791–92, 794, 799, 800–801, 1148–49
- *inductive-statistical, 676, 691, 706–16, 720, 725, 733, 750–52, 759, 764, 767, 770, 774, 775, 777, 779–84, 786, 790–91, 793, 799, 800–801, 1148–49, 1260. *See also* Explanation(s), probabilistic
- inference to best. *See* Inference, to best explanation
- and irrelevance objections, 676–77, 720–25, 727–28, 730–31, 784–88, 800
- of laws, 686–87, 696, 720–21, 742n9, 771, 774, 785, 906, 908, 910–11, 976
- laws in, 39–40, 42–43, 48, 50, 55, 76–77, 675, 678–82, 685–90, 692, 694n2, 695, 697, 702–3, 716, 722–23, 727, 729, 731, 733–36, 739–41, 747–50, 754, 757, 761, 767–71, 777, 779–80, 785, 789–90, 794, 796, 799, 800–801, 805, 806–7, 829–30, 836–37, 846, 869, 871, 875–76, 896–97, 945n6, 987, 1108
- narrative, 1149–51, 1260–61, 1282
- *ontic conception of, 802
- partial, 692–94, 736, 773
- and prediction, 40, 129, 133–34, 228, 232–34, 611, 676, 695–703, 724, 742n15, 747–48, 762–63, 769, 773–79, 782, 792, 798, 799–800
- probabilistic, 688–90, 693, 698, 709–10, 746, 748–49, 754, 756–58, 763–64, 782, 791–93, 798, 801–2. *See also* Determinism, and probabilistic explanation; Explanation(s), deductive-statistical; Explanation(s), D-N-P model of; Explanation(s), inductive-statistical
- and problem of ambiguity, 676, 709–11, 714–15, 718n3, n7, 750, 780–83, 791–92, 800, 803n10
- and properties, 737–39, 789–90
- and reduction. *See* Reduction(s), and explanation
- reductive, 903, 906, 908–10, 912, 915, 926, 972, 1005–6, 1011, 1016–17, 1023–25, 1030, 1032, 1041–42, 1044. *See also* Reduction(s), and explanation
- and specificity requirement of maximal (RMS) 676–77, 712–16, 718n3, n5, n7, 750–51, 758–60, 780–81, 783–84, 786–87, 791–93, 800
- role of, in science, 1241, 1242
- of scientific beliefs, 393–95
- self-evidencing, 699–701, 776–77
- single-statement view of, 677, 732, 735, 788–90, 800
- sketches, 693–94, 773
- statistical-relevance view of, 1149–51, 1260, 1282

- Explanation(s) (*cont.*)
 and *structural identity thesis, 676, 695–703, 724, 762, 769, 773–79, 782, 790, 799
 and *symmetry objection, 720, 725–28, 773–74, 787–88, 790–91, 799, 1106, 1113n14
 of theories, 926, 950, 953, 1007, 1109
 and total evidence requirement. *See* Evidence, total, requirement of
 and truth, 134, 462–63, 525–27, 751, 759, 764, 770–72, 780, 867, 1101–2, 1106–7, 1109, 1129–31, 1149, 1188, 1190–92, 1202, 1205n5, 1268
 ultimate, 1102–6
 *Explanationism, 411, 487–89, 514–15, 525–33, 535–36, 542
 and historical thesis of evidence, 411, 481–82, 485
 Explanatory power, 328, 384, 388–89, 395, 400, 585, 598, 721, 800, 807, 837–38, 869, 903, 1000n20
 as criterion for theory assessment, 5, 32, 98, 129–30, 133–34, 212, 225–26, 232–34, 383, 386, 654–55, 767, 1066, 1077, 1106, 1109, 1118, 1233, 1237
 and factual content, 869, 896
 and predictive accuracy, 104
 *Extension(s), 839, 886–87, 893, 914, 1010, 1037. *See also* Meaning(s), and extension (reference)
 Extensionality, 839, 843n5, 886–87
 Fallibilism, 152–53, 237
 about knowledge, 323
 Falsifiability, 6–7, 15, 221–22, 255, 307, 327–28, 353n49, 398, 511–12, 813
 criterion of, 2, 7–9, 22–23, 27, 29, 30–31, 40, 43–44, 46n9, 48–49, 51, 52, 59, 63–69, 71, 72, 76–78, 505, 527. *See also* Testability
 logical vs. prescriptive/methodological, 64–65
 of physical laws, 303
 and simplicity, 410, 456–57, 522–23
 Falsification, 24, 79, 185, 247, 261, 303, 355, 366, 431, 463, 527, 660, 936. *See also* Reduction ad absurdum
 *ambiguity of, 195, 246, 261–63, 273, 277, 302–304, 309, 326, 327, 356–57, 366, 374, 376, 382, 397–98, 582n27
 and evidence, 463
 of generalizations by counterexamples, 435
 impossibility of, 30, 221–22, 255, 323, 326, 356, 359–65, 376, 398
 rule of, Popperian, 326, 327, 332, 335, 352n49, 386–87
 saving of theories from, *See* Theories, retention of, in the face of evidence
 *Falsificationism, 410, 430–31, 463, 505, 507–10, 540, 542, 549, 1187. *See also* Falsifiability, criterion of
 Feminism, 199, 206, 247–48, 250, 253n17. *See also* Empiricism, feminist; Epistemology, standpoint; Postmodernism, feminist; Science, feminist critiques of
 Fertility. *See* Fruitfulness
 Field theory, 65
 Flood, Biblical, 41, 49, 50–51, 58–59, 75, 77, 1126
 Forces, 867–71, 873, 896–97, 899, 901n11, 940, 1069, 1159, 1227
 Formalism. *See* Confirmation, and formalism
 Freudianism. *See* Psychoanalysis
 Fruitfulness, 103, 114, 117, 120, 125, 129–30, 135, 157, 182, 230, 232–34, 247, 251n8, 339, 391, 395, 552, 576–77, 640, 645–46, 668, 1128
 Gases, kinetic theory of, 911–12, 914, 957, 972, 997n7, 1005, 1008, 1010
 Gender, 200, 206, 247
 bias in science. *See* Bias
 differences in intelligence, 192, 194, 196, 246
 ideology, 192–93, 195, 197, 203–4

- Generality, 211–12
- Generalizations, 69, 185, 212, 216,
382, 736, 739–40, 749, 814,
848, 1080, 1182. *See also* Hy-
potheses; Law(s)
inductive, preferences among, 435–
42
- Genes, Mendelian, 1175, 1181,
1251–52, 1255, 1266
- Genetics, 51, 247, 253n17, 690, 779,
797, 847, 971, 972, 993, 1005,
1029, 1181, 1251–52
classical, 903–4, 971–92, 994,
1028–38, 1042–44
Mendelian, 20, 26, 40, 971–80,
1028, 1251–52, 1255. *See also*
Law(s), Mendel's
molecular, 971–80, 982, 985–96,
998n9, 1028–38, 1042–44,
1251–52, 1266
population, 39
- Geometry, 222, 381
analytic or synthetic?, 368–70
Euclidean, 315–17, 364, 370, 377,
520
non-Euclidean, 316–17, 370, 377,
520
pure vs. empirical (interpreted),
370, 403n13
reduction to absurdity in, 264,
266
- Gestalt switch, 119, 121, 139, 141,
144, 154, 218, 227. *See also*
Conversion
- Good sense, 278–79, 309–11, 314,
316, 364, 375–76, 381, 398
- Gravitation/Gravity, 29, 51, 449n,
559, 561n, 586, 680, 796,
1016, 1061, 1103–4, 1113n13,
1121, 1159, 1176
Cartesian theory of, 392
explanation of, 96–97, 99, 151n,
225, 1102–4
law of. *See* Law(s), of gravity
theory of,
Einstein's, 4–7, 8, 99, 251n7,
599, 1194. *See also* Relativity
Newton's, 10n4, 20, 23–24, 63,
67, 68, 69, 70, 71, 225, 266–
70, 303–4, 308, 357, 363–64,
375, 392, 398, 447, 451–52,
457, 461–62, 513, 517, 531,
565, 584, 599, 655, 687, 908–
10, 929, 1007–8, 1015, 1040,
1046n15, 1091–92, 1102–4
- Grue. *See* Induction, new riddle of
- Heat, 114, 682, 1126
caloric theory of, 90, 91n, 1126,
1141–42, 1173, 1255
kinetic theory of, 113, 682, 906–7,
911, 927, 935–36, 1005–6,
1013, 1023, 1045n12, 1119–
21, 1166
- Heuristics, 24, 216–17, 579, 950,
953, 956, 960, 967n5, 1024,
1053
- Historical context of theories, 2, 32,
33–34, 135
- Historical thesis of evidence, *see* evi-
dence, historical thesis of
- *Holism, 171, 185, 204, 330, 561n.,
647, 1122. *See also* Duhem-
Quine thesis; Duhem thesis;
Paradigm shifts, holistic view
of; Paradigm(s), holistic char-
acter of
and Duhem-Quine thesis, 255, 295–
96, 313, 315, 317, 346–47,
350n19, 356–58, 360–61,
363–64, 366, 374, 381, 383,
386, 397–400, 401n2, 402n8
and egalitarian thesis, 325, 340
and Quine thesis, 404n17
scope of, 374–76, 398
and theory-ladenness of observation,
358
and underdetermination, 348n9
- Homology, 43, 76
- Hydrogen
atom, Bohr's theory of, 452, 516,
951
emission spectrum of, 450–51, 452,
512, 516
- *Hypothesis, 185, 263, 273, 355, 426,
510–11. *See also* Theory/
Theories
ad hoc, 7, 43, 74, 129, 133–35,
278, 362, 549, 568, 571, 587,
593, 611, 621, 646, 1130, 1181
auxiliary, 7, 24, 25, 71, 195, 203–4,
261–62, 303–7, 327–30, 356–
58, 360–62, 365, 375–77, 383,
388–89, 398–400, 553, 568,
570–71, 577–78, 582n27, 698,
823, 837, 1142, 1149

- *Hypothesis (*cont.*)
 - catchall, 556, 562, 566, 568–70, 572–73, 578, 643–44, 657
 - confirmation of. *See* Confirmation; Evidence
 - conflict of, with established theory, 452–53
 - corroboration of. *See* Corroboration
 - credibility of, 452, 457–58, 523–24
 - fundamental, how refuted. *See* Testability, of fundamental hypotheses
 - genetic, 983–84, 1035, 1043
 - high-level vs. low-level, 382, 675, 916, 1182, 1266
 - isolated. *See* Testability, of isolated hypotheses
 - Newton's rejection of, 21, 69, 82n20, 150–51, 266, 941, 1103–4, 1146
 - predictive content of, 434, 437, 442
 - probability of, 457–58, 464–65, 471
 - testing of. *See* Testing/Tests, of theories
 - theoretical support of, 451–53
- *Hypothetico-deductive model, 46n8, 135, 151, 332, 341n27, 410, 507, 525, 549, 553, 638–40, 655, 672n19, 1145–46, 1259. *See also* Confirmation, hypothetico-deductive; Explanation, hypothetico-deductive; Induction, hypothetico-deductive model of
- Idealism, 561n, 678–79, 905, 1055, 1091, 1095, 1111n3, 1112n7, 1166, 1174, 1198–99, 1219–21, 1229, 1270, 1276
- Idealization, 124, 964, 970n36, 1025, 1171
- *Identities/Identification(s), 858, 957–60, 974, 1026–27, 1039, 1042
 - empirical discovery of, 891, 914–15, 1011
- Impartiality, and sociology of science, 393
- *Incommensurability
 - and meaning variance, 117, 916–20, 929–32, 939–41, 1156
 - of paradigms/theories, 117, 119–122, 125, 131–32, 135, 139, 144, 155, 159, 167, 185, 222–24, 231, 575, 580, 916–20, 929–32, 968n12, 1156, 1172–73
 - and realism, 131–32, 135, 1201
- Indeterminism. *See* Determinism/Indeterminism
- Individualism, 158, 173, 175, 186, 193, 205, 240, 248
- *Induction, 4, 63, 112–14, 128, 150–51, 184, 232, 266–69, 283, 362, 409–10, 412–28, 431, 437, 451, 495–506, 516, 538–39, 576, 652–53, 689–90, 694, 695, 697, 707, 711, 750–51, 763–64, 1179, 1259. *See also* Confirmation; Evidence; Inference, probabilistic/statistical; Logic, inductive
 - and arguments from analogy, 495–96, 538, 564, 576
 - and causal inference, 1176–77, 1264, 1283
 - causal model of, 502
 - and circularity, 414–19, 427–28, 497–500, 505, 539
 - conservative (“More of the Same”), 416–17, 418, 420, 423, 499–500, 501–2, 542. *See also* Induction, enumerative; Induction, straight rule of
 - deductive invalidity of, 498–99, 538
 - and discovery, 409–10
 - eliminative, 151, 504, 531, 663
 - and entrenchment, 390
 - enumerative, 151, 391
 - and explanation, 767, 791–92
 - to generalizations, 363–64, 409, 495–96, 498, 1189, 1268
 - Humean problem of, 128, 390, 410, 415–18, 427, 497–501, 505–6, 539, 848–49, 1145
 - hypothetico-deductive model of, 422–23, 502–3. *See also* Confirmation, hypothetico-deductive; Hypothetico-deductive model
 - instantial model of, 420–22, 501–3
 - new riddle of, 336–38, 389–91, 400–401, 421, 422, 405n28, 503, 544n8, 550, 611, 617, 849
 - Newton's method of, 266–70
 - paradoxes of, 336

- to a particular, 495, 497–500, 653
 Popper's rejection of, 406n31, 410, 426–28, 434, 505–6, 509, 514, 522, 527, 640
 pragmatic problem of, 434, 440
 pragmatic vindication of, 520, 539, 542
 principles of, 413–23, 427–28, 495–97, 499, 505–6, 538–39, 1176, 1264
 problem of. *See* Induction, and problem of justification; Induction, Humean problem of
 and problem of description, 410, 412, 414, 417, 418–25, 495–97, 501–5, 538–39
 and problem of justification, 339, 390, 410, 412, 414–19, 425, 426–28, 434, 439, 495–501, 505–6, 521, 538–39, 618, 942, 1015, 1145, 1147, 1201–2, 1259, 1264, 1268
 revolutionary (“Time for a Change”), 416–17, 418, 420, 500
 and statistical syllogism, 495, 498, 538
 straight rule of, 336, 520–21, 542. *See also* Induction, conservative; Induction, enumerative
 and uniformity of nature, 416, 443, 499, 543n3, 506
 Inductive logic. *See* Logic, inductive
 Inductive support. *See* Confirmation
 *Inductivism/Anti-inductivism, 150–51, 355, 363, 398, 411n1, 431n5, 434, 442, 443, 505–10, 523, 527, 542, 549, 941
 Inertia, principle of. *See* Law(s), of inertia
 Inference, 323, 412–14, 510, 579, 592, 620, 635, 648–49, 1169, 1203, 1211, 1263. *See also* Law(s), as rules of inference; Logic; Methods/Methodology
 ampliative, 323, 325–26, 336, 383–84, 386, 390, 409, 495, 540, 1146, 1259–60. *See also* Induction; Method(s)/Methodology, ampliative
 *to best explanation (abduction), 462, 498–99, 502, 525–26, 543n2, 1050, 1075–77, 1095–96, 1164, 1177, 1179, 1188, 1203, 1236–37, 1242, 1259, 1264, 1271, 1279. *See also* Realism, and inference to best explanation
 causal, 423–24, 503–5
 deductive. *See* Deduction; Logic
 inductive. *See* Induction
 nondeductive. *See* Inference, ampliative
 principles/rules of, 413–14, 415, 417–18, 584, 1108, 1199, 1236–37, 1279. *See also* Induction, principles of; Law(s); Logic
 probabilistic/statistical, 549, 559, 587, 666, 703, 750–52, 1162, 1177, 1190
 Information. *See* Disconfirmation
 *Instrumentalism, 84, 132–35, 233–34, 580n2, 614, 867, 911–12, 924n, 1008, 1052, 1054, 1057, 1067–68, 1077–78, 1088–89, 1100, 1126, 1130, 1155, 1166–67, 1187, 1191–92, 1194–95, 1198–99, 1209–10, 1214–16, 1221, 1228–29, 1233, 1262, 1269–70, 1272, 1286n4. *See also* Law(s), instrumentalist
 view of; Realism; Reduction(s), instrumentalist account of
 Instruments. *See* Experimental apparatus
 Integrity, intellectual, 41, 45, 211
 *Intensions, 839, 886–87, 893, 1010. *See also* Meaning(s)
 Intentionality, 839, 843n5, 886–87
 Intuitionism. *See* Logic, intuitionist
 Irrationalism/Irrationality, 84, 102, 113, 119, 122, 141–44, 222, 224, 227, 239, 242, 249–50, 552–53, 1172. *See also* Rationality
 Irrelevant conjunction. *See* Confirmation, and tacking paradox
 Judgment, in theory choice, 115–16, 132, 134, 278, 552, 563, 574, 579. *See also* Good sense
 Justification, 154, 411n2, 1150–51, 1265
 *context of, 107–9, 113, 172, 193, 200, 203, 240, 248, 429, 502, 504, 539, 544n7, 579, 638, 1054

Justification (*cont.*)

- of deduction, 414, 500–501
- and hypothetico-deductive model, 410, 638

Kepler's laws. *See* Law(s), Kepler's

Kinds

- in creation-science, 43, 58–60, 75
- *natural, 856–57, 892–96, 898, 979, 1033, 1047n26, 1155, 1180–82, 1212–13, 1262, 1265–66, 1273, 1283

Knowledge, 20–22, 129, 323, 419

- a priori, 368–71, 373–74, 381, 398–99, 403n16, 404n17, n23, 428, 856, 879, 891, 1220n
- logic of vs. psychology of, 242, 428–30

- total body of, 296, 313–14, 317, 374, 472–73, 710, 780

Lamarckianism, 153, 183, 252n14

Language(s), 117, 120, 135, 292, 295, 312, 366, 370–71, 981, 1066, 1082, 1083, 1174, 1216, 1227, 1233

- acquisition, 412–13, 419, 420, 496–97

- artificial, 283, 289–92, 373, 399, 458, 524, 557n, 617, 640

- extensional, 287–88

- natural, 1068

- observation. *See* Observation, language

- ordinary, 945n3

- physical-object, 1056, 1059–61, 1062n5, 1063n9

- sense-datum, 293–95, 1055–56, 1062n5

*Law(s), 185, 211, 257, 409, 510–11, 678–84, 805–901, 1032, 1182–83, 1190, 1199. *See also* Generalizations; Hypotheses; Necessity, of laws; Regularity/Regularities

- approximate, 909–11, 913, 920, 974, 1008, 1016, 1040, 1078, 1171

- Bode's, 1101

- Boyle-Charles, 113, 652, 688, 726, 741, 773, 986, 997n7, 1036

- of bivalence, 378

- *bridge. *See* Reduction(s), and bridge laws

- causal, 382, 723, 727, 729, 742n15, 747–48, 761, 770, 791, 811, 814, 816, 829, 862

- chemical, 97n

- and confirmation, 21, 617, 624n14, 821, 827, 829, 831, 833, 847–48, 884

- contingency of, 891, 893, 898, 899

- Coulomb's, 868–71, 874–75, 896

- and counterfactuals. *See* Conditionals, subjunctive/counterfactual as definitions, 271, 813

- and demarcation problem, 39–40, 41–42, 48, 50–51, 55–58, 76, 79

- descriptive, 257, 267, 269, 276, 363, 382, 389, 767, 1018, 1054, 1167n3. *See also* Law(s), empirical; Regularity/Regularities

- deterministic. *See* Law(s), universal

- distribution, 380, 404n20, n23

- of double negation, 378–79

- empirical, 675, 678, 682, 686–87, 768, 771, 811, 814, 893, 914, 920, 1010–11, 1039, 1080, 1227, 1238. *See also* Law(s), descriptive; Regularity/Regularities

- of equality of action and reaction, 96, 275

- of excluded middle, 297, 317, 378–79, 380–81, 400, 849, 1223n1

- experimental. *See* descriptive; empirical

- and explanation, 39–40, 42–43, 48, 50–51, 55–56, 76–77, 675–77, 678–82, 685–90, 694n2, 695–97

- explanatory, 382, 865–69, 875–76, 884, 896–97

- facticity of, 865–71, 873, 875–76

- of falling bodies, 271–72, 447, 451–52, 455, 517, 574–75, 686–88, 908–9, 929, 972, 997n7, 1007–8, 1016, 1023, 1038, 1040

- falsifiability of, 270–77, 306–8, 313, 813

- falsity of, 865–66, 868, 875, 896–97, 1288n26

- *functional, 770–71, 818, 824, 850, 853, 882–84

- functionalist theory of, 829–30, 843n6, 846, 1036
 fundamental, 865–66, 875, 897
 Galileo's. *See* Law(s), of falling bodies
 in genetics, 973–80, 987–88, 1030–32, 1036
 of gravity, 23, 39, 71, 77, 151n 267, 268, 303, 308, 357, 475, 574–75, 576, 687, 771, 867–71, 874–75, 896, 899, 982. *See also* Gravitation/Gravity
 Hardy-Weinberg, 39
 Hooke's, 726, 770, 882–83
 Hubble's, 882
 of ideal gases, 882–83, 951, 972–73, 1010
 and identification problem, 887–88, 899
 of inertia, 95n, 274–75, 364, 479, 881–82, 939–40, 1014
 and inference problem, 887–89, 899
 instrumentalist view of, 807n1, 911–12, 1008–9, 1171
 Kepler's, 21, 39, 40, 266–70, 363–64, 382, 398, 574–75, 594, 598–99, 655, 656, 686–88, 771, 908–10, 929, 991, 1007, 1015, 1038, 1040
 Leavitt-Shapley, 445–46
 logical, revision of, 297–317, 326, 377–81, 386, 404n23
 Mendel's, 40, 55, 974–77, 1031–33, 1181, 1266
 and missing-values problem, 882–84
 of **modus ponens*, 326, 342, 376, 500–501, 717n2, 755, 795, 1076
 of **modus tollens*, 303, 326, 327, 356, 386, 397, 507
 of motion, 271, 945n6
 Newton's, 23, 39, 40, 96, 303, 306, 314, 357, 364, 382, 398, 506, 574–75, 686–88, 771, 812–13, 817–18, 847, 850, 861, 881–82, 896–97, 939–40, 960, 1023, 1038, 1040. *See also* Gravitation, theory of, Newton's; Mechanics, classical (Newtonian)
 of multiple proportions, 275–76
 *necessitarian theory of, 805–6, 809–15, 846, 850, 852, 855, 885–87, 891, 898–99. *See also* Necessity, of laws
 of noncontradiction, 381, 1223n1
 *noninstantial, 806, 817–18, 881–82, 890, 898
 observational. *See* Law(s), descriptive; Law(s), empirical
 Ohm's, 726, 844n7
 opacity of, 828–30, 839, 886–87
 Planck's radiation, 382, 951
 probabilistic, 683, 688–90, 746, 749, 752–55, 758, 761, 763, 779, 793–96, 801
 psychological, 683, 773, 865n, 907
 of radioactive decay, 752–55
 of rational indices, 276
 of rectilinear propagation of light. *See* Light
 and reduction, 906–14, 916–17, 919, 972–80, 1005–6, 1008, 1015, 1023, 1030, 1037, 1039, 1042
 *regularity theory of, 805–6, 815–24, 879–85, 890–96, 897, 898–99
 *epistemic, 806, 822–24, 829–38, 846, 883–85, 898
 simple, 806, 816–21, 826–29, 880–83, 1036
 as rules of inference, 820–21, 911–12, 1008
 Snell's, 313, 382, 446–48, 457, 511, 688, 770
 statistical, 676, 683–84, 688–90, 703, 706–8, 712–14, 716, 749–50, 751, 767, 775, 779–81, 790, 794, 808–9, 815–16, 847, 849, 1013, 1045n12, 1078
 super, 875–75
 symbolic, 269–70
 theoretical, 1018, 1080, 1155–56, 1167n3, 1171, 1184n3, 1238
 of thermal expansion, 683
 of thermodynamics, 67, 747n, 798, 808, 832, 841, 865, 896, 929, 935–36, 1013, 1036
 ultimate vs. derivative, 882
 universal, 683–84, 688–90, 703, 754–55, 777, 790, 805–6, 807n1, 815–16, 865n, 1013, 1045n12, 1078, 1182, 1238–29, 1266
 and *universal generalizations, 806,

- *Law(s) (*cont.*)
 818–23, 827–43, 844n7, 846,
 848, 862, 865, 880–81, 883–
 85, 888, 898–99, 1036. *See*
also Law(s), *regularity theory
 of
 universals theory of, 806, 830–31,
 838–43, 859–62, 885–90,
 891, 898–99
 unknown, 824, 884–85, 898
 *vacuous, 816–18, 850–53, 860–
 61, 880–81. *See also* (Law(s),
 noninstantial
 Wiedemann-Franz, 957–59
 Length contraction, 13n, 178, 571–
 72, 1019–20
 Lexicography, 283
 Leyden jar, 97, 123, 1121
 Light, 13n, 39, 537, 547n39, 1121,
 1168n10. *See also* Ether theo-
 ries; Experiment(s), on polar-
 ized light; Experiment(s), on
 velocity of light; Optics
 electromagnetic theory of, 98–99,
 453, 847, 957, 967n2, 1005,
 1011
 gravitational bending of, 4n, 6, 25,
 63, 166, 529–31, 603, 606n20
 particle (emission) theory of, 79n,
 94, 153, 154, 262–63, 265–66,
 278, 305–6, 314, 359–60,
 361–62, 365, 528–29, 569,
 571–72
 quanta, theory of, 453
 rectilinear propagation of, 273, 571
 refraction of, 263, 360, 453, 686,
 1102–3, 1121
 selective absorption of, 572, 645
 velocity of, 263, 265, 278, 305, 310,
 359–60, 362, 365, 572, 747n,
 847, 854
 wave theory of, 94, 98–99, 153,
 154, 265–66, 278, 305–306,
 309–10, 359–60, 361–62,
 365, 453, 528, 566, 569, 571–
 72, 577, 767, 803n5, 929,
 1119, 1124
 Logic, 171, 212, 296, 328, 344, 367,
 369, 371, 373–74, 375, 379,
 381, 398–400, 404n17, n18,
 n23, 495–96, 499, 607, 666,
 683, 810–15, 891, 1082
 Aristotelian, 311, 317
 changes in, 297, 317, 326–27, 342,
 379–81, 386
 classical, 378–81, 404n23, 1203,
 1211
 of confirmation, 135, 517
 deductive. *See* Deduction
 of discovery, 409–10, 426, 507,
 544n7, 545n14
 holism regarding, 315, 317, 376,
 378–81
 inductive, 22, 334, 337, 326n, 426–
 28, 431, 451, 510, 516, 523–
 24, 556, 585, 617, 620, 703,
 711–12, 717n2, 718n3, 782,
 1118
 intuitionist, 317, 379–80
 of justification, 409–10
 laws of, 282, 296, 297, 343, 369,
 381, 404n23, 636, 699, 754,
 810. For particular laws, *See*
 Law(s)
 many-valued, 378
 modern, 311, 378, 585, 588
 of prediction, 435
 of probability, 683
 quantum, 317, 380–81, 404n20,
 n23
 and rationality. *See* Rationality, and
 logic
 of reduction, 1005–11, 1285n1
 role in scientific inference, 309–10,
 314, 320, 323, 343, 362, 375–
 76, 381, 384, 393, 395, 397,
 398
 and social factors, 343
 truths of, 282–83, 311, 369, 371,
 373, 378, 615, 810, 856
 Logical empiricism, *See* Empiricism,
 logical
 Logical positivism, *See* Positivism,
 logical
 Logicism, 369–70
 Lysenko episode, 20, 183, 200, 244
 Magic, 29, 72
 Magnetism, 561n, 680–82, 768–69,
 899
 Marxism, 4–5, 8–9, 15, 23–25, 34,
 63, 64, 219
 Mathematics, 14, 114, 124, 257, 273,
 294, 296, 298, 360–62, 367–
 71, 375, 379, 398, 404n17,
 691, 771, 772, 803n7, 1082,

1214. *See also* Arithmetic;
Geometry
as analytic, 369–70, 373–74,
403n16, 404n18
a priori knowledge of, 368–69,
856
foundations of, 1187–88, 1267–68
holism regarding, 315–17, 376–77,
381
objects of, 379–80
as synthetic, 368
- Meaning(s), 173, 221–22, 280, 292,
293, 296, 366, 371, 373–74,
375, 398–99, 867, 944, 1022,
1155–56, 1173, 1227–28,
1262, 1265, 1282. *See also*
Observation, terms, meaning
of; Reference; Synonymy; The-
oretical terms, meaning of
and analytic truths. *See* Analytic
statements; Analytic-synthetic
distinction
descriptivist theory of, 1172
and essence, 281
experimental, 274–77, 307
and extension (reference), 223, 281,
285, 288, 893, 919–20, 1021–
22
and holism, 350n19, 398–99, 904
and intension (*sense), 223, 281,
886, 893, 1021–22
and naming, 281, 1155
physical, 276, 277
of theoretical terms, *See* Theoreti-
cal terms, meaning of
theory-dependence of, 120, 217–
18, 222–24, 904, 916–19,
925–26, 929–32, 939, 1009,
1018–22, 1040–41, 1071,
1094, 1155–56, 1172, 1234
variance of, 220, 222–24, 238, 249,
812, 917, 929–32, 938, 939,
940–41, 955, 1014–15, 1017–
20, 1058, 1155–56, 1172. *See
also* Incommensurability, and
meaning variance; Reduc-
tion(s), meaning-invariance
condition on; Theoretical
terms, meaning of
*verifiability criterion of, 371, 373,
399, 427n, 505, 1221, 1285n1
verification theory of, 292–93, 295,
300n15, 367
- Meaningfulness, 9, 274–77, 300n15,
371, 399, 1227–28, 1234,
1285n1
- Measurement, 124, 136n, 313, 520,
913, 1049, 1074, 1095, 1236
- Mechanico-corpuscularianism, 95n,
96–97
- Mechanics, 114, 267–68, 270, 271–
72, 274–75, 796, 870, 874,
897, 1001n21, 1025, 1042
celestial, 24–25, 123, 151, 270,
316, 341, 916, 1102
classical (Newtonian), 23–25, 67,
68, 103, 123, 303–304, 306–7,
309, 311, 316, 341, 364, 506,
567, 574–75, 577, 584, 651,
655, 741, 774, 906, 908–10,
913, 929–30, 951–53, 955–
57, 959–60, 962, 964, 968n6,
972, 991, 997n7, 1007–8,
1015, 1017, 1019–27, 1040–
41, 1046n15, n20, 1091–92,
1124
fluid, 797
Newtonian and Cartesian, testing
of, 341–42, 392–93
quantum, 24, 50, 89, 99, 103,
115–16, 154, 177, 297, 317,
364, 380, 404n23, 427n, 452,
690, 746, 755, 762, 796–98,
802, 874–75, 905n, 906, 911,
917, 924, 927–28, 936–39,
945n6, 947n26, 948n30, 951,
956, 959–60, 1005, 1007,
1026–27, 1052n, 1058, 1081,
1105–6, 1110, 1112n8, 1149,
1158, 1163, 1186, 1194–96,
1207n25, 1221, 1239–40,
1269, 1288n30. *See also* Revo-
lution, quantum
relativistic, 1017. *See also* Relativity
statistical, 690, 903, 912–13, 919–
20, 959, 965, 967n5, 997n7,
1010, 1013, 1045n12,
1047n27, 1239. *See also* Ther-
modynamics, and statistical
mechanics
- Mechanism(s), 34, 430n, 747–49,
755, 761, 763, 791, 794–95,
801–2, 996n4, 1033–34
- Medicine, 16, 29, 67, 72, 74, 264,
1126
norms and values in, 197–98
- Men, 194–95, 196, 249
as human norm, 197–99

Men (*cont.*)

- skeletal ideal of, 197–98
- Mercury, 133, 233, 819
 - anomalous perihelion of, 166, 304, 308, 513, 529–31, 599, 601, 656, 658–59
- *Metaphysics, 8–9, 14, 21, 65–66, 95, 99, 177, 179–80, 188n25, 216, 280, 292, 312, 371, 678–79, 767–68, 809, 830, 903, 922, 924–26, 937, 941–42, 944, 945n4, 1011, 1014–15, 1077, 1090–91, 1093–94, 1106, 1110, 1112n5, 1139, 1170–71, 1180, 1196, 1216, 1221–22, 1254, 1273, 1275–76, 1285n1, 1288n33
- Meteorology, 16, 67, 797
- *Method/Methodology, scientific, 95, 99, 170–71, 173, 174, 179–80, 181, 185, 199, 201–2, 205, 212, 217, 226–27, 235–38, 239, 246–50, 255–56, 261, 320–44, 419, 404n17, 439, 507, 520–21, 585–88, 606n11, 608, 614–15, 618–21, 642, 678, 698, 922, 932, 981, 1012, 1028, 1039–40, 1054, 1081, 1096, 1106, 1139, 1172, 1177, 1187, 1199, 1202, 1240, 1244, 1270–71. *See also* Criteria for theory choice; Falsificationism; Hypothetico-deductive model; Inference; Objectivity, of methods; Values
 - ambiguity of, 156–60, 237–39
 - ampliative, 323–24, 326, 333, 336, 338, 342–43, 346, 351n27, 386, 584, 1146. *See also* Induction; Inference, ampliative
 - Bayesian explanations of, 549–50, 587, 592–93, 607, 615–16, 646, 650–51, 656, 659, 663, 666, 669
 - change of, 124, 140–56, 204
 - contextualist account of, 173, 240
 - critiques of, 320–22, 323–24, 330–31, 335, 346
 - culture-bound, 204
 - disagreement about, 162
 - experimental, 258–59, 264, 355, 361, 363, 1154
 - ideal organon of, 333–34
 - inconsistency of, 160–61
 - of inference, 323
 - Kuhn's critique of, 139, 141, 156–67
 - Mill's, 161, 423–24, 502, 503–5
 - Newtonian, 21, 124, 150–51, 161, 186n1, 266–70, 363, 1103–4, 1146
 - pluralist, 923–25, 933–39, 942, 943–44, 948n30, 1011–15, 1040
 - positivist account of, 171, 185–86, 240
 - and underdetermination. *See* Underdetermination, of theories by evidence/methods
- Miracle(s), 39, 41, 45, 56, 76
 - success of science as, 1099, 1110, 1111n3, 1136–38, 1140–41, 1151, 1240, 1245, 1264, 1279, 1283
- Moon, 28, 451, 473
 - motion of, 24, 71, 134, 447
 - origin of, 125
- Motion, 178, 593, 1194
 - inertial, 474
 - natural, 96
 - Newton's laws of. *See* Laws, of motion
- Myth, 5, 8, 63, 298, 678, 938–39, 948n33, 1014
- *Natural kinds. *See* Kinds, *natural
- *Natural ontological attitude (NOA), 1200–1204, 1204n1, 1207n27, 1208n29, 1209–13, 1215–16, 1218, 1221, 1223, 1267, 1270–77, 1283–85
- *Necessity, 287–88, 372, 707. *See also* law(s); rigid destination
 - causal, 806, 879–80
 - Hume's view of, 806, 811–14, 879–80, 888
 - of laws, 805–6, 807n1, 809–15, 821, 827, 838–41, 846, 850, 855–56, 866, 880, 887–89, 891–96, 898–99, 1182, 1266. *See also* Law(s), necessitarian theory of
 - logical, 810–15, 855, 879–80, 888, 891
 - metaphysical, 806, 814–15, 855–56, 891–92, 898–99
- Negotiation, 407n39
- Neptune, 28, 31, 583n41

- discovery of, 4n, 71, 270, 304, 461–62, 463
- postulation of, 71, 304, 461, 577
- Nonrealism. *See* Antirealism; Natural ontological attitude (NOA)
- Nominalism, 842–43, 862, 890, 1082
- Normal science. *See* Science, normal
- Objectivity, 83, 107, 115–16, 170–86, 193, 200, 205–6, 212–14, 228, 239, 241–45, 246–50, 400, 439, 552–53, 574, 577–79, 917, 1173, 1194, 1216, 1218–19
 - and background assumptions, 85, 179–80, 241–42
 - and criticism, 85, 178, 181, 183–85, 214, 243–45, 247, 250
 - degrees of, 181, 185, 250
 - diversity required for, 85, 175–76, 179, 181, 184–85, 188n33, n34, 205, 243, 249, 923–24, 939. *See also* Methods/Methodology, pluralist
 - and empirical method, 220
 - and evidence, 176, 179
 - as freedom from bias, 116, 170, 180, 245
 - as intersubjective agreement, 227
 - of methods, 170–71, 173, 201, 239–40
 - and publicity, 176, 179
 - as truth, 170, 239, 245, 247
- *Observables/Unobservables, 39, 150–51, 212, 409, 420–22, 424, 502–3, 767–69, 847, 867, 912, 917, 1008, 1018, 1049–51, 1053–62, 1068–69, 1071–77, 1079–80, 1089, 1094–99, 1102, 1109, 1112n7, n8, 1122, 1126, 1128–29, 1146, 1153–54, 1157, 1195, 1207n27, 1210, 1221, 1227–31, 1234–44, 1259–66, 1268, 1277–80, 1282. *See also* Observational-theoretical distinction; Theoretical entities
- Observation(s), 4, 11, 43, 63, 69, 100, 220, 245–46, 355, 922, 932, 944, 1011, 1018, 1049, 1054–55, 1061, 1072–73, 1089, 1094–98, 1142–43, 1145–46, 1157, 1218, 1221, 1227–31, 1235, 1243–44
 - and confirmation, 8–9, 29, 1060, 1278
 - direct, 1055, 1057–58, 1061, 1095
 - and falsification, 30, 397
 - language, 176, 221–22, 241, 358, 367, 503, 917, 925, 931–32, 1018–20, 1055–57, 1059–61, 1062n5, 1063n9, 1071, 1094, 1231, 1234. *See also* Observation, statements; Observation, terms
 - of physical objects, 241
 - statements, 29, 176, 241, 303, 375, 397, 426, 598, 654, 911–12, 916–18, 925, 1008–9, 1054, 1060–61, 1142, 1228
 - terms, 912–13, 1044n4, 1057, 1059–60, 1073, 1098, 1115, 1227–29, 1231, 1244–45, 1277. *See also* Theoretical terms
 - meaning of, 185, 221, 916–19, 925, 931–32, 940, 1009, 1018–22, 1040–41
 - *theory-ladenness of, 15n, 176, 195, 197, 217–18, 220–22, 238, 246, 249, 258–59, 357–58, 912–13, 916, 919, 931–32, 934–36, 938, 1020–21, 1040–41, 1071, 1094, 1142, 1171, 1234
- Observational-theoretical distinction, 30, 358, 614, 912, 931–32, 1018–20, 1044n4, 1055–62, 1062n5, 1071–75, 1094–98, 1227–31, 1241–45, 1277–78. *See also* Observables/Unobservables
- arbitrariness of, 1056, 1061–62, 1072, 1278
- ontological significance of, 1050, 1056, 1062, 1062n5, 1074, 1095, 1227–28, 1230–31, 1235–36, 1278
- terminological vs. ontological, 1227, 1234, 1243, 1245
- as theoretical, 1056, 1278
- Occult qualities, 95–96
- Oedipus effect, 9n3
- *Ontology/ontologies, 294, 297, 298, 925, 950, 1137, 1200–1202, 1270–71. *See also* Reduction(s), ontological
- of paradigms/theories, 131–32, 140,

Ontology/ontologies (*cont.*)

142–46, 150–51, 162, 216–17, 231, 233, 235, 673n22, 916, 931, 1017, 1115, 1190

and science, 298

*Operationalism, 366–67, 940, 1020, 1228

Optics, 14, 89, 262–63, 359–60, 361–62, 571–73, 906, 911, 914, 929. *See also* Ether theories; Light

Cartesian, 340

and electromagnetic theory, 915, 951, 953, 955–56, 958–59, 1005, 1011, 1023, 1025–6

Newton's vs. Huygens', 265–66, 306

Oxygen theory of combustion, 12n, 67, 86, 94, 103–4, 108, 122, 123, 143–44, 231

*Paradigm(s), 33, 39, 72, 78, 83, 86–100, 140–56, 159, 173, 215–19, 244, 249–50, 552, 1034.

See also Paradigm shifts; Rationality; Research programmes, scientific; Revolutions, scientific; Theory/Theories

change of. *See* Paradigm shifts

change within, 142, 144–55

choice between. *See* Paradigm shift

circularity in the justification of, 88, 100, 121–22, 126–27, 142, 148, 219

components of, 123, 124, 140–42, 144–50, 216, 235, 250

debates between, 124–25, 142–156, 162, 163, 218, 224–27, 230, 239, 250, 278

as disciplinary matrices, 124, 216

as exemplars, 216, 413, 496–97

holistic character of, 217, 236, 239

incommensurability of. *See* Incommensurability, of paradigms/theories

normative functions of, 95–100, 127;

rival, 88–100, 121–22, 141, 149, 156–58, 162, 163, 167, 219, 222–26, 238, 250, 338, 392

Paradigm shifts, 86–100, 102, 122–27, 131–32, 140–56, 165, 217–19, 237–38, 339, 1200,

1271. *See also* Conversion;

Revolutions, scientific

and crisis, 86–87, 93, 123, 136n1

holist view of, 126, 132, 141–46, 148, 150, 151–55, 218

rationality of. *See* Rationality, of paradigm shifts

resistance to, 102, 120, 123, 157

uni- vs. multi-traditional, 148–50

Parapsychology, 1

Parity, 449, 1160, 1165, 1263

Peer review, 175, 181, 214

Personalism, 532, 546, 558–62, 564, 578, 582n28, 588, 590, 624n16, 634, 640–42. *See also* Bayesian confirmation theory, subjective

tempered, 549, 561–62, 593, 641

Persuasion, 87–88, 102–3, 122, 219, 552

*Phenomenalism, 366–67, 1052, 1054–56, 1060–61, 1111n3, 1198–99, 1215–16, 1221, 1229, 1231, 1242, 1270, 1278

Phlogiston theory of combustion, 12n, 67, 92, 94, 97n, 98, 103–4, 108, 123, 143–44, 165, 231, 458n5, 771, 1118, 1120–21, 1126, 1141–42, 1173, 1180–81, 1255, 1266

Physical objects, 297–98, 1020, 1056, 1062n5, 1216, 1275

phenomenalist reduction of. *See* Phenomenalism

Physics, 5, 16, 33, 39, 44, 50, 51, 97, 114, 125, 150, 154, 158, 187n5, 192, 206, 237, 296, 297, 308, 317, 366, 374–77, 380–81, 394, 404n23, 431n6, 537, 563, 573, 677n1, 755, 779, 802n3, 801, 813, 865–76, 897, 905n, 909, 928, 936–37, 946n22, 962–63, 971–80, 991, 1028, 1041, 1049, 1061, 1074, 1077, 1095, 1097, 1121–22, 1153–55, 1158, 1159–60, 1164–66, 1167n3, 1171, 1182, 1184n3, 1194–97, 1203–4, 1221n, 1236, 1244, 1261, 1265, 1266. *See also* Dynamics; Mechanics; Reality

Aristotelian/scholastic, 95–97, 123–24, 127, 131, 158, 178, 228–29, 232, 340, 932, 944n2

- holistic character of, 263–64, 355–56
- Newtonian, 23–25, 34, 39, 71, 91–92, 95, 95n, 123–24, 146, 150, 175, 303–4, 340, 576, 577, 924n, 928–29, 932, 1069. *See also* Dynamics; Gravitation/gravity; Law(s), of gravity; Mechanics
- testing of theories in, 257–79, 302–8, 313–14, 355–62, 376, 381, 397–98, 400, 401n2
- Philosophy, 257–60, 355, 374, 401n2, 680, 1121, 1126
- experiments in, 258–59, 313, 376
- relation to fundamental sciences, 259–60, 376, 993, 996
- Planets, 71, 72
- influence of, 10n4, 27–28, 30, 32
- Kepler's theory of orbits of, 21, 39, 40, 221–22, 463, 517, 655.
- See also* Law(s), Kepler's
- motions of, 21, 24, 39, 40, 133–34, 229, 233, 267–70, 303–304, 357, 363–64, 398, 447, 655, 686–87, 818, 1102. *See also* Law(s), Kepler's
- Plate tectonics, 29, 40, 51, 1119, 1128, 1182
- Poisson bright spot, 528–29, 533, 565–66, 572, 577, 1121
- Popperianism, 514–15. *See also* Falsificationism
- *Positivism, logical 34, 91, 135, 167n2, 171–73, 185, 206, 240, 366–67, 369–71, 373, 375, 399, 505, 519, 580n2, 1052n, 1059, 1064, 1067–68, 1070–71, 1088–91, 1094, 1105, 1110, 1122, 1157, 1166, 1186, 1206n15, 1209–10, 1214, 1227–28, 1272, 1285n1
- *Possible worlds, 852–57, 862
- Postmodernism, feminist, 200, 206, 248, 250–51
- Postulates, 138n, 291, 950, 952
- Pragmatism, 280, 299, 314, 375, 1093–94, 1129, 1198–99, 1210, 1270
- Precession of equinoxes, 29, 31–33, 72
- Prediction(s), 2, 15–17, 28, 31, 42–43, 65, 68, 104, 211–12, 384, 386, 423, 433–43, 460–61, 507–8, 516, 682–84, 711, 798, 805, 829–30, 833, 837, 843n6, 846, 848, 871, 884, 911, 1049, 1110, 1153, 1227, 1240
- and confirmation, 7, 220, 513–15, 525, 527–31, 536, 541, 618, 621. *See also* Predictionism/Predictivism
- and evidence, 460–67, 471, 481–82, 486–88, 490–91, 525–31, 534–35, 543, 611. *See also* Predictionism/Predictivism
- and explanation. *See* Explanation(s), and prediction
- generalizations used in making, 435–22
- *novel, 24–25, 71–72, 103, 159, 166, 225, 229, 238, 239, 339, 391, 402n5, 410–11, 450–51, 460–63, 481–82, 486, 512–16, 518, 525–31, 533–36, 540–41, 549, 565, 577, 599, 611, 642, 1100–1101, 1110, 1130, 1137, 1144, 1147–48, 1187, 1191, 1202, 1246, 1253–54, 1257, 1281
- and practical decisions, 410, 434, 436–43, 509, 540, 684
- rational, 434–43, 509–10, 520, 540
- reasons for making, 433–34, 509
- and tests, 30, 40, 63–64, 78, 769, 799
- and theory assessment. *See* Accuracy, predictive
- *Predictionism/Predictivism, 410–11, 466, 471, 478n17, 481–82, 485, 487–88, 514–15, 518, 525–33, 535–36, 541–42, 669
- Predictive capacity. *See* Accuracy, predictive
- Primatology, 194, 203
- Probabilism, 22, 152–53, 598, 732–33, 746
- Probabilities/Probability, 22, 69–70, 134, 427n, 428, 471, 516, 556–58, 585–87, 589–90, 598, 608, 676, 707, 713, 717n2, 718n5, 780, 798. *See also* Bayesian confirmation theory; Bayes's Theorem
- axioms of, 458, 471, 523, 557, 588–89, 594, 621n1, 628–29, 635, 636, 647, 657–59, 667, 669

Probabilities/Probability (*cont.*)

- Bayesian view of. *See* Bayesian confirmation theory
 - conditional, 629–30, 657, 667
 - and confirmation, 390, 517, 523–24, 541, 550, 585–86, 598, 782
 - and degrees of belief. *See* Belief, degrees of
 - epistemic, 761, 763–64, 765n8
 - and frequencies, 559–60, 562, 581n9, 564, 578, 585–86, 610–11, 622n4, 616, 640–42, 689, 763, 797, 801
 - of hypotheses, 457–58
 - logical, 458, 524, 556–57, 560, 585–86, 616–17, 640, 689–90, 703, 711, 715
 - objective, 557–58, 559–60, 562, 566, 578, 610, 616, 624n16, 622n4, 640–42, 643, 648, 761, 764, 793, 796, 802
 - personal, 557–58, 561, 563, 566, 578, 583n42, 1067. *See also* Bayesian confirmation theory; Personalism; Probabilities/Probability, subjective
 - physical, 748, 753–58, 761, 765n8, 796–97, 801
 - posterior. *See* Bayesian confirmation theory, and posterior probabilities
 - prior. *See* Bayesian confirmation theory, and prior probabilities
 - and propensities, 559–60, 761, 796–98, 801–2
 - statistical, 688–89, 706–8, 712–14, 718n5, 753, 761, 765n8
 - subjective, 610, 613, 616, 623n10, 628, 634, 640–41, 643, 648, 649, 656, 665, 667, 752. *See also* Probabilities/Probability, personal
 - theory of, 70, 458, 464, 471, 524, 549, 557, 585, 592, 607, 609, 613, 615, 617, 619, 627, 628, 667, 714, 783, 784, 816, 903
- Problem(s), scientific, 162, 904, 981–85, 1034
- solving, 31–33, 89–90, 102, 139, 165, 173, 224, 228, 234, 987–88, 990, 1034–35, 1036, 1043. *See also* Puzzle solving
 - standards for solutions to, 95–100, 123, 161–62, 163

weighting of, 100, 163–67, 203, 220, 224–25, 238

- Progress, 23–25, 27, 27n, 32–33, 71–74, 102, 187n4, 202, 216–17, 219–20, 222, 224, 227, 230, 249, 903, 922–23, 927, 933–34, 941, 943, 1005, 1007, 1011–14, 1028, 1115, 1137, 1141–44, 1147, 1200–1201, 1281, 1271. *See also* Research programmes, scientific
 - and cumulativeness, 216–17, 227, 903
 - and objectivity, 903
 - and realism, 83, 131, 1115–16, 1131, 1194, 1246–47, 1253–54, 1256, 1269
- Proof. *See* Demonstration
- Propensities. *See* Probability/probabilities, and propensities
- Pseudoscience(s), 1–2, 3–4, 9, 10n4, 15, 20, 22, 24–26, 27, 29, 30, 32–35, 64–65, 67–69, 71–74, 76–79, 192, 505. *See also* Demarcation of science from pseudoscience; Research programmes, scientific
- Psychoanalysis, 1, 4–6, 8, 15, 16, 24, 34, 39, 63, 146, 563, 582n24, 584, 692, 773
- Psychology, 8, 32, 72, 83, 140, 146, 183, 502, 506, 607, 614–15, 618, 773, 865n, 907, 1005, 1061
- individual (Adlerian), 4–6, 8, 63
 - neuro-, 179
- Publicity, 176–77
- and objectivity, 179
- Puzzle solving, 11–12, 14–17, 33, 67–68, 84, 93, 124–25, 131–32, 215–16, 215n2, 224, 228, 233. *See also* Problem(s), scientific, solving
- Quantum mechanics. *See* Mechanics, quantum
- Quine thesis, 255, 302, 311, 313–14, 317, 374. *See also* Duhem-Quine thesis; Duhem thesis; Holism
- Radiation, 87
- Rationality, 33–34, 206, 214, 327, 439, 577–79, 594, 598, 1067. *See also* Objectivity

- and Bayesian confirmation theory, 589, 613, 615, 634, 635–38, 640, 647–48, 650, 667
- of cognitive values, 126, 147–48, 167n2, 225–26, 228, 230–33, 236–37
- as consensus, 227, 552–53
- of disagreement, 112, 668
- *hierarchical model of, 84, 141, 144, 148, 235–37, 250
- and logic, 278, 320, 327, 330, 343, 397, 636, 658
- of paradigm shifts, 87–88, 102–3, 119, 121, 125, 127, 141–48, 155, 157, 162, 163, 219, 222, 224–27, 230, 338–39
- pragmatic, 327
- *reticulation model of, 84, 140, 144, 149–50, 152, 236–37, 250
- of science, 84, 119, 128, 130–32, 141–42, 158, 159, 172–73, 180, 184, 192–93, 200–201, 204, 219–20, 227–28, 235, 241, 325, 331, 342–46, 400, 439, 478n12, 508, 549, 552–53, 556, 574, 640
- of scientific community, 205
- of scientific revolutions, 23, 25–26, 71, 83–84, 102–3, 117–18, 119, 122, 125, 219, 222, 224, 227, 230, 233–34, 239, 249–50
- scope of, 205
- of theory choice, 117–18, 119–22, 125–28, 131, 135, 141, 145–46, 155, 159–60, 163–65, 200–203, 212, 226, 234, 324–31, 334, 384, 385–89, 397, 399, 400, 439–40, 508–10, 552–53, 1139, 1254. *See also* Criteria for theory choice; Theory/Theories, choice; Values
- Rational reconstruction, 429, 507, 544n7, 691
- Raven paradox. *See* Confirmation, and raven paradox
- *Realism, 84, 131–36, 170, 230, 233–34, 235, 250, 252n9, 442–43, 580, 621, 859, 861–62, 866–67, 875, 1049–53, 1064–71, 1074–76, 1079, 1082–84, 1088–94, 1096, 1099–1111, 1111n3, 1112n7, 1114–32, 1132n3, n4, 1137–39, 1140–41, 1146–48, 1153–55, 1166–83, 1186–1204, 1204n1, 1209–23, 1227–30, 1232–43, 1245–56, 1259, 1263–85, 1287n17. *See also* Antirealism; Empiricism; Instrumentalism
- abductive arguments for. *See* Realism, and inference to best explanation
- and Bayesian confirmation theory, 614, 654
- convergent, 1115–16, 1126, 1131, 1140, 1142, 1166, 1246–52, 1270, 1284
- entity, 1051, 1065, 1146, 1153–55, 1157–58, 1167, 1170–76, 1180–83, 1260, 1261–67, 1282–83
- and epistemic values, 135
- experimental, 1051, 1158, 1169–83, 1253, 1261–67, 1282
- and independent world, 176–77, 1066, 1082, 1216–22, 1275–77, 1284, 1288n33
- and inference to best explanation, 498, 1050, 1075, 1116, 1129–30, 1164, 1169, 1178, 1183, 1188–92, 1197, 1205n7, 1236–37, 1247, 1265, 1267–69, 1279–80, 1282–83
- and language, 1083
- about laws, 806, 897
- metaphysical, 206, 1087n27
- and natural ontological attitude (NOA), 1051
- and quantum mechanics, 1083, 1186, 1194–96
- and relativity theory, 1194–95
- and success of science, 131, 1050–51, 1076–77, 1083–84, 1099, 1114–32, 1133n12, 1134n15, 1136–51, 1158, 1169, 1176–78, 1182, 1187, 1191–92, 1194, 1196, 1200, 1206n8, 1228, 1240, 1245–56, 1260–61, 1263–65, 1267–72, 1279–83
- about theories, 1153–55, 1170–73, 1261–62, 1264, 1282–83
- transcendental, 867
- “ultimate argument” for, 1082–84,

Realism (*cont.*)

- 1099–1101, 1240, 1245–46, 1279–80
- and underdetermination, 255, 321, 1049
- Reason, 83, 173
- Reasonableness, 320, 330, 362, 365, 375, 381, 398
- *Reductio ad absurdum, 261, 264, 266, 360–61, 398, 854
- Reduction(s), 903–1044, 1050, 1068, 1239
 - and appearance-reality distinction, 906, 1005
 - and application, 957, 1026
 - approximative, 955, 960–62, 1016, 1040
 - and *bridge laws, 904, 913–15, 916, 920, 930, 939, 954–55, 957–59, 972–73, 975–79, 997n8, 1010–11, 1012, 1014, 1021–22, 1024–27, 1030–32, 1039–42
 - and conceptual refinement, 989, 1036–37, 1043
 - *consistency condition on, 926–39, 944n2, 954, 955–56, 968n14, 1006, 1012–17, 1021, 1023–26, 1039–41
 - correspondence account of, 912–16, 920, 1009, 1015–17, 1022–26, 1039, 1044n6, 1285n1
 - derivational. *See* Reduction(s), and logical derivability
 - domain-combining (reduction₁), 950, 952–56, 961–62, 1024–26, 1041
 - domain-preserving (reduction₂), 92n, 950, 953–57, 960–66, 1024–26, 1028, 1041
 - and explanation, 675, 903, 906–7, 953, 959, 972, 990, 993–94, 1006, 1009, 1023–24, 1030, 1039, 1041–42. *See also* Explanation(s), reductive
 - and explanatory extension, 990–91, 993–94, 1036, 1037–38, 1043–44
 - heterogeneous. *See* Reduction(s), inhomogeneous
 - *homogeneous, 903–4, 907–10, 911, 954–55, 958, 1007–8, 1015, 1024–25, 1039, 1041, 1044n6
 - *inhomogeneous, 903–4, 907–8, 910–16, 954–55, 1007–11, 1024–25, 1039, 1041
 - instrumentalist account of, 911–12, 1008–9, 1023
 - and logical derivability, 223, 904, 908–10, 913–14, 916, 920, 926, 950, 953–57, 959–63, 966, 968n17, 972–73, 1006–10, 1012, 1016, 1020, 1023–17, 1030, 1038–39, 1041–42, 1044, 1044n6, 1286n8
 - logic of. *See* Logic, of reduction
 - *meaning-invariance condition on, 904, 912, 926–32, 939–41, 943, 955–56, 1006, 1012, 1014–15, 1017–22, 1025, 1039–41, 1045n14
 - ontological, 950, 952, 952–55, 959–60, 1023–25, 1027, 1041
 - partial, 953, 956, 963
 - and presupposition, 1036–37, 1043
 - and replacement, 908–10, 915–20, 1016, 1017–18, 1020–21, 1023, 1040
 - and scientific change, 517
 - and symmetry problem, 1027–28
- *Reductionism, 293, 295, 366–67, 371, 829, 835, 1054, 1203
- and analytic-synthetic distinction, 295–96, 367, 373, 399
- in genetics, 972–80, 988–96, 1000n20, 1029–38, 1042–44
- phenomenalist. *See* Phenomenalism
- radical, 293–96
- Reference, 223–24, 281, 285, 828n, 117, 1255
- new (causal) theory of, 857–58, 892, 893–96, 898, 1155–56, 1172–73, 1173, 1249, 1262, 1265, 1282–83
- and realism, 132, 170, 867, 1050, 1066, 1084, 1112n10, 1115–23, 1173–74, 1177, 1180–81, 1200, 1216, 1218, 1255, 1262, 1266, 1271, 1280–81, 1287n12, n17. *See also* Reference, and success of science
- stability of, 132, 1115–16, 1123, 1132n5, 1140, 1142, 1155–56, 1180, 1200–1201, 1246–47,

- 1250, 1255–56, 1266, 1271, 1284
- and success of science, 1084, 1112n10, 1115–23, 1128–29, 1131, 1140–41, 1177, 1246–53, 1255, 1271, 1280–81
- and truth, 170, 1126, 1128, 1140, 1200, 1211, 1247–48, 1250–52, 1255, 1281
- Reflexivity, and the strong programme in sociology of science, 393, 394
- Refutability/refutation, *see* falsification; falsifiability; revisability; testability
- Regularity/Regularities, 39, 41, 1078–80, 1084, 1102, 1104–6, 1171, 1176–77, 1238–39, 1245, 1288n26. *See also* Law(s), descriptive; Law(s), empirical
- *Relativism, 144, 169n38, 206, 219, 322, 384, 385, 917, 1219, 1276, 1285, 1289n35
 - cultural, 34
 - Kuhnian arguments for, 84, 219–26, 238
 - about truth, 84, 1223n1
 - and underdetermination, 255, 321–22, 337, 346, 383
- Relativity,
 - general theory of, 4–6, 8, 13, 24, 63, 67, 99, 251n7, 304, 308–10, 317, 364, 370, 377, 449n, 481, 512–13, 529–31, 576, 601, 603, 656, 658–59, 687, 865, 905n, 916, 945n6, 1194, 1269
 - special theory of, 13n, 154, 178, 213, 316, 376, 512, 520, 572, 577, 928–31, 946n22, 949n3, 951–53, 955–56, 960, 962, 964, 968n6, 1017, 1019–23, 1025, 1040–41, 1046n18, n20, 1092, 1194, 1269
- Reproduction, role of egg and sperm in, 193–94, 203, 245
- *Research programmes, scientific, 65, 78, 152, 179, 1069
 - methodology of, 23–26, 71–72, 364
 - progressive vs. degenerating, 24–26, 33, 71–72, 78
- Resemblance thinking, 74
- Retrograde motion, 105, 134, 229, 233
- Revisability,
 - of all statements, 297, 374
 - of conventional hypotheses, 273–74, 306–8
 - of Euclidean geometry, 315–17, 377
 - of logic, 317, 377–81
 - of mathematics, 377
 - of theories, 40, 44, 48, 49–51, 56, 58–59, 76–77, 273–74, 306–8
- Revolution, chemical, 86, 97, 98, 99, 114–15, 123, 142–43, 145, 165–66, 231
- Revolution, Copernican, 34, 67, 68, 84, 86, 122, 123–24, 129, 132–35, 225–26, 229–33, 250, 949n43. *See also* Astronomy
- Revolution, quantum, 124, 231, 364, 949n43
- Revolution, Scientific, 127, 154, 1183
- *Revolutions, scientific, 12–14, 17, 26, 57, 67–68, 84, 86–100, 120–27, 136n1, 150–52, 154–55, 216–19, 224, 249–50, 251n6, 308, 364, 458n5. *See also* Paradigm(s); Paradigm shifts; Rationality, of scientific revolutions; Science, extraordinary
 - compared with political, 86–88
 - conceptual transformation in, 93–94, 231, 932, 1200. *See also* Meaning, theory-dependence of; Theoretical terms, meaning of
 - consensus in, 240
 - depth of, 121–25, 126–27, 131, 231
 - holistic, 141–56
 - necessity of, 91
 - not cumulative, 86, 89–90, 217
 - rationality of. *See* Rationality, of scientific revolutions
- *Rigid designator, 858, 892
- Rule(s), 1075–77. *See also* Criteria of theory choice; Inference, principles/rules of; Law(s), as rules of inference; Method(s)/Methodology; Values
 - of correspondence. *See* Reduction(s), and bridge laws

Rule(s) (*cont.*)

- hidden, 412–14, 496–97
- for theory choice. *See* Criteria of theory choice; Methods/Methodology
- semantical, 289–92, 399

Science(s), 127–28, 137n21, 170, 211, 232, 280, 295, 296, 298, 321, 608, 684, 1093, 1221, 1232

- ancient, 14
- vs. crafts, 16, 67
- as cumulative, 83, 89–90, 99
- essence of, 48, 52, 77, 228, 231–32
- an experience, 39
- experimental, 257, 374
- and explanation, 39, 767, 1079–81, 1084, 1099, 1101–10
- extraordinary (revolutionary), 12–15, 67–69, 93, 100, 215, 218–19, 224; *See also* Revolutions, scientific
- feminist critiques of, 192–94, 199–200, 205–207, 214
- goals of, 129–31, 174, 235–37, 250
- growth of, 12, 34, 922, 927, 991
- historical, 66, 677n1
- history of, 26, 66, 69, 71, 83, 89, 99, 130, 142, 148, 171, 185, 213, 217–19, 239, 240, 304, 305, 474, 502, 529–30, 579, 585, 607, 945n4, 946n23, 1114, 1126–27, 1140, 1153, 1158, 1191, 1201, 1247–49, 1255, 1259
- mature/immature, 124, 150, 215, 1084–85, 1115–17, 1122, 1127–28, 1140, 1248, 1251, 1255
- Newtonian. *See* Physics, Newtonian
- *normal, 11–14, 27, 33, 67–68, 87, 88, 89–90, 93, 95, 100, 123, 125–26, 163, 215–19, 231, 244. *See also* Puzzle solving
- objectivity of. *See* Objectivity and philosophy, 8, 14, 39, 618, 665–67, 670
- philosophy of, 52, 66–67, 83, 106, 107–9, 112–13, 119, 135, 139, 144, 171, 184, 192, 200, 248, 255, 345, 397, 551–52, 579–80, 608, 618–21, 665–67, 670,

- 903, 906, 923, 946n23, 1011, 1028–29, 1070, 1101, 1153–54, 1158, 1166, 1183, 1202–3, 1221, 1271–72, 1285n1
- plausibility considerations in, 558–59, 571–74, 576, 577–79, 639, 642, 645–46
- as practices, 981, 1034–35, 1043
- publicity of, 176–177, 179
- and religion, 20, 26, 38–41, 54, 56
- social, 192, 203, 677n1, 801, 813, 906
- social character of, 174–80, 181, 186, 189n35, 193, 214, 239–41, 242–43, 247, 248–50
- social influences on, 82, 84, 118n8, 144, 158, 164, 171–72, 174–80, 192–93, 195, 197, 200–207, 213–14, 227, 250, 256, 321, 342–44, 393–97, 400, 401, 478n12, 584, 917. *See also* Bias; Values, contextual
- sociology of. *See* Sociology of science
- subjective factors in, 102–3, 105–6, 109–10, 112, 121, 156–60, 164–67, 171–73, 175, 178–80, 181, 185–86, 200–205, 213–14, 226–28, 238–39, 240, 241–43, 245, 247, 250, 258–59, 279, 321, 338–39, 345, 478n12, 553, 636, 670, 945n4. *See also* Bias; Subjectivism; Values, contextual
- success of, 561–62, 641–42, 1154, 1195, 1202, 1206n8, 1233, 1253, 1281. *See also* Realism, and success of science; Reference, and success of science; Truth, and success of science
- as tool for prediction, 297
- as unit of empirical significance, 296, 314, 374–76, 399
- values in. *See* Values
- Scientism, 608, 620–21, 665, 670
- Scope, 1021, 1144
 - as criterion for theory assessment, 103–4, 111–12, 114, 125, 157, 229, 247, 251n8, 339, 391, 401, 552, 574–75, 618, 645
- *Sense data, 293–95, 297, 1052, 1055, 1059n, 1061, 1066, 1082, 1112n7, 1218, 1229,

- 1231, 1276, 1278. *See also*
 Language, sense-datum
- Set theory, 294, 370, 1187, 1267–68
- Simplicity, 351n27, n32, 333, 389,
 400, 459n8, 515–16, 576,
 585–86, 612, 663, 622n7,
 673n22, 1111n4, n5, 1142,
 1147
- Bayesian explications of, 549–50,
 576, 593, 650–54, 665, 669
- criteria of, 229–30, 454–55
- as criterion for theory assessment,
 103, 105, 110–11, 120, 128–
 29, 132, 134, 157, 158,
 168n12, 173, 212, 225–26,
 229–30, 232–33, 235, 247–
 48, 338, 339, 345, 349n15,
 386, 391, 401, 410, 453–58,
 518–24, 540–42, 552, 562–
 63, 618, 640, 642, 645–46,
 650–54, 668–69, 1066, 1092–
 94, 1233
- and curve fitting, 456, 518, 520–
 21, 594–96, 651, 670
- descriptive vs. inductive, 520–21
- and economy, 455, 519–20, 542,
 1227
- and empirical content, 456–57
- and falsifiability. *See* Falsifiability,
 and simplicity
- of fundamental laws, 519
- as goal of science, 41, 298–99
- and justification, 455–57
- and *likelihoods, 653–54
- and prior probabilities, 550, 612–
 13, 663, 670
- and truth, 134, 212, 516, 518, 522,
 541–42, 1092, 1094
- Skepticism, 255, 944, 1075, 1089,
 1129–30, 1157, 1173, 1223,
 1241
- Cartesian, 415
- deductive, 501, 539
- inductive (Humean), 410, 414–19,
 439, 497–501, 505–6, 539,
 879, 1145–47, 1260
- Sociobiology, 194
- Sociology, 83, 618, 797
- of science, 83, 121, 171, 176, 185,
 321, 343, 353n54, 502, 584,
 917n, 1192
- *strong programme in, 393–97,
 401, 406n35, n36, 917n
- Species, 65, 697–98, 775, 892,
 900n10, 1175
- mutability of, 49, 75, 698
- Standpoint epistemologies. *See* Episte-
 mology/Epistemologies,
 standpoint
- Stellar parallax, 567–68, 570–71, 645
- Strong force/interaction, 1159
- Structural identity, thesis of. *See* Ex-
 planation(s), and structural
 identity thesis
- Subjectivism, 84, 125, 136n1, 144,
 153, 172, 227, 230, 239, 242,
 549, 641, 667–68, 752, 880,
 888, 898
- Subjectivity, 84, 115–16, 171–73,
 175, 177, 181, 214
- Symmetry, 563–64, 642
- and the strong programme in soci-
 ology of science, 393–94
- Symmetry thesis. *See* Explanation(s),
 and structural identity thesis
- Synonymy, 1045n14
- and analyticity. *See* Analytic-
 synthetic distinction
- arguments from, 942–43, 1015
- cognitive, 286–88, 292
- and definition, 283–85
- and interchangeability, 285–88,
 372
- and verification theory of meaning,
 292–93, 399
- *Synthetic statements, 289, 292, 295,
 312, 366–68, 371, 373, 399,
 404n17, 427–28
- a priori, 368–69, 371, 428, 506
- depend on fact, 280, 371
- *Tautology, 65–66, 1067. *See also*
 Logic, truths of
- and analytic truth, 369–70, 72,
 399, 404n22, 427
- Tentativeness. *See* Revisability
- Testability, 30, 40, 43, 48–49, 51,
 64–66, 72, 76–78, 235, 306–
 308, 522, 611, 676. *See also*
 Falsifiability; Falsification
- of creation-science, 43–44, 49–51,
 58, 76–77
- of fundamental hypotheses, 275–
 77, 362
- of isolated hypotheses, 49, 263,
 270, 276, 279n11, 302–3, 306,
 313–14, 315

Testability (*cont.*)

of theories, 7, 8, 15, 398, 916

Testing/Test(s), 63, 68–69, 79, 123,
165–66, 185, 377, 436, 456,
463, 511, 770–71, 821, 1049,
1166

an alternative theories, 31, 265–66,
305, 361, 612, 917, 933–34,
936–37, 1013

and confirmation, 7, 258, 445

deductive, 428, 507

of personal conjectures, 11–12, 15

of physical theories. *See* Physics,
testing of theories in

severe, 536–37

of theories, 7, 11–15, 17, 165, 179,
257–79, 302–8, 313–14,
350n19, 355–62, 364–65,
376–77, 383, 390, 392, 397–
98, 429–31, 434, 439–40, 483,
489, 540, 767

Theology, 21–22, 38, 83, 212, 371,
809, 922, 1093, 1203, 1285n1

Theoretical entities, 150, 910, 1052,
1054–55, 1060–61, 1063n9,
1065, 1069, 1071, 1096–97,
1122, 1132n3, 1146, 1150,
1153–58, 1169–71, 1175–83,
1184n3, 1200, 1209–10, 1229,
1232, 1234–40, 1261, 1264,
1277. *See also* Observables/
Unobservables; Observational-
theoretical distinction

Theoretical statements, 598, 654,
1049–50, 1054, 1094,
1133n13, 1216, 1228. *See also*
Law(s), theoretical

Theoretical terms, 358, 912, 1050,
1056, 1071, 1115, 1227–29,
1231, 1234, 1276

elimination of, 1228

meaning of, 117–18, 120, 185,
217, 222–24, 269–70, 271–
72, 904, 912, 916–19, 926–32,
940, 1009, 1014, 1018–22,
1040–41, 1054, 1057, 1068,
1155–56, 1228, 1234, 1262,
1265, 1282

*Theory/Theories, 21, 31, 39, 90,
185, 211, 212, 298, 322, 355,
367, 383, 409, 426, 739–41,
910, 916, 1005–6, 1034, 1039,
1043, 1049–51, 1054, 1064,
1089, 1119, 1142, 1153–54,

1166, 1169, 1182, 1184n3,
1227, 1232, 1256. *See also* Hy-
potheses; Paradigm(s); Para-
digm shifts; Research
programmes, scientific; Sci-
ence; Underdetermination

acceptance of, 3, 9, 172–73, 176,
185, 627, 667, 1066–67,
1069–70, 1074, 1076, 1089,
1092, 1198, 1209–10, 1221,
1223, 1228, 1232–34, 1241,
1270, 1272, 1278, 1283–84

aesthetic appeal of, 127, 132–34,
233, 1139, 1254

choice, 103–18, 119–36, 136n1,
226, 323–42, 386, 468, 470,
472, 474, 520, 553, 555–56,
562, 577, 1090–94, 1139,
1172, 1241. *See also* Criteria
of theory choice; Methods/
Methodology; Paradigm(s),
change within; Paradigm shifts;
Rationality, of theory choice;
Science, social influences on;
Science, subjective factors in;
Underdetermination, of theo-
ries by evidence/methods

Bayesian algorithms for, 109,
556, 558, 570, 573–74, 579,
640, 644–45, 668, 672n14

comparative, 201–3, 331, 569,
574, 578, 643–44

and empirical adequacy, 331

conflict between, 90–94, 264–66
comparability of, 120, 159, 167n2,
917, 938, 1021, 1040–41,
1172–73. *See also*

Incommensurability

competition among. *See* Theory/
Theories, rival

conservatism, in change of, 299,
925, 965, 1189–92

empirical adequacy of. *See* Empiri-
cal adequacy

empirical support for, 325, 329–32,
335, 337, 339, 346–47,
349n10, 350n22, 385, 388–89,
400, 420–22. *See also* Confir-
mation; Evidence; induction
empirically equivalent, 1069, 1090–
94, 1105

epistemic (truthlike) character of,
119, 131, 135, 170

and experiments, 257–66, 269,

- 271–74, 355, 1177, 1183,
1263. *See also* Experiment(s),
Theory-dependence of
falsification of. *See* Falsifiability;
Falsification
generation of, 200–203, 205, 214,
250, 411n1, 502, 507, 538. *See also* Discovery, context of
instrumentalist view of. *See*
Instrumentalism
microstructural, 912, 915, 928,
1078, 1080, 1238–40
proof of, 69–70, 324
range of application of, 92–93
retention of, in the face of evi-
dence, 7, 8, 23, 25, 30–31, 40,
43–44, 56, 58–59, 67, 71, 72,
78, 195–96, 246, 272–74,
277–78, 303n, 314, 317, 321,
323, 324–31, 334–35, 339–
40, 344–45, 349n13, 350n19,
353n50, 356, 362, 364–65,
374–76, 381, 385–89, 396–
97, 398–401, 404n17, 453
rival, 14, 32, 72–73, 84, 124–25,
202–3, 220, 324, 331, 335,
359, 361, 384, 391–92, 398,
401, 489, 512, 531, 536–37,
575, 578, 601, 634, 644, 650,
663–64, 767, 917, 1189
simplicity of. *See* Simplicity
structure of, 46n9, 980, 997n6,
1005, 1028, 1031
testing of. *See* Test(s)/Testing
theoretical support for, 451–53, 510,
540–41, 576
truth of, 3, 5, 9, 1049–51, 1065–
68, 1070, 1077–78, 1089–90,
1092, 1101–2, 1105, 1109,
1115–16, 1120, 1124, 1130,
1136–38, 1140–44, 1154,
1169–70, 1179, 1182–83,
1188–89, 1196, 1228, 1232–
33, 1246–48, 1255, 1261,
1277, 1285. *See also* Truth
underdetermination of. *See* Under-
determination, of theories by
evidence/methods
usefulness of, 1077–78, 1237–38
Thermodynamics, 67, 89, 90–91,
747n, 798, 911–14, 927,
947n29, 987, 1001n21, 1036,
1047n26, 1124. *See also*
Law(s), of thermodynamics
and Brownian motion, 164, 934–
36, 1013
and statistical mechanics, 903, 912–
13, 919–20, 929, 936, 951,
959, 967n5, 1005–11, 1023–
26, 1041, 1068
Translation, 117–18, 120–21
Transubstantiation, 40–41, 55–56
Truth, 170, 182, 230, 239, 255, 290–
92, 396, 414, 428, 641, 867,
937–39, 1008–9, 1088–89,
1094, 1096, 1109, 1112n7,
1117, 1129, 1142, 1192, 1198–
1200, 1203, 1209–17, 1223n1,
1232–34, 1268, 1270, 1272–
74, 1284, 1287n12, n17. *See also* Analytic statements; Syn-
thetic statements; *Theory/
Theories, truth of
approximate, 131, 219, 227, 230,
352n49, 539–40, 1009, 1049–
50, 1067, 1084, 1115–16,
1120, 1122–26, 1128–29,
1133n10, 1137, 1140, 1141–
45, 1147, 1169–70, 1177,
1179, 1182, 1188–93, 1196,
1199–1200, 1202–3, 1206n11,
1246–48, 1250–52, 1253,
1255, 1256–60, 1264, 1268,
1280–81, 1284
as correspondence, 1189, 1199–
1200, 1203, 1212–13, 1216–
17, 1223, 1268, 1270–71,
1273, 1284, 1288n33
depends on language and on fact,
292, 295, 312, 367, 370–71
and explanation. *See* Explana-
tion(s), and truth
and inductive inference, 497–99
logical. *See* Logic, truths of
as rationalconsensus, 245, 1207n27,
1208n29
and realism. *See* Realism
and reference. *See* Reference, and
truth
relative, 1143, 1256–57
and simplicity. *See* Simplicity, and
truth
and success of science, 1120,
1124–26, 1128, 1131, 1137–
39, 1147–50, 1177, 1246–48,
1250–57, 1259, 1260, 1280–82
Tarskian theory of, 1203, 1211–16,

Truth (*cont.*)

1271–74, 1276, 1284,
1288n30

as value in science, 211–13, 232,
1066, 1083, 1170

values, 378

*Underdetermination, 320–47, 366,
402n9. *See also* Duhem-Quine
thesis; Holism

ampliative, 324–35, 335–42, 344,
346, 382, 385, 400, 413

causal, 156, 343–46, 397

and deconstructionism, 321

deductive. *See* Underdetermination,
Humean

descriptive vs. normative, 325, 334,
344, 346, 386, 400

and egalitarian thesis (EGAL), 323,
324–25, 331–32, 334–39,
346, 347, 348n9, 352n36, n37,
353n59, 385–86, 389, 390–91,
400–401

global, 236–38, 296–98, 313–14,
391

Humean (HUD), 323–26, 34, 342–
43, 346, 350n22, 352n37,
383–85, 395, 400, 401

and induction, 412–16, 419–20,
425, 495–98

local, 156, 160, 237–39, 391–93,
401

and nonuniqueness thesis (NUT),
324–25, 334–38, 346, 347,
348n9, 349n11, 352n37, n39,
353n59, 385, 389

overestimated, 202, 322

Quinean (QUD), 328–30, 334–35,
342, 349n14, 350n19, n22,
351n23, 366, 387–89, 399

and rational retention of theories,
321, 323, 324–31, 334–35,
339–40, 344–45, 349n13,
350n19, 353n50, 386–89, 397,
400–401

and realism. *See* Realism, and
underdetermination

and relativism. *See* Relativism, and
underdetermination

of theories by evidence/methods,
153–56, 156–67, 195, 197,
202, 246, 255–56, 323–47,
350n22, 351n27, n35, 382–97,
399–401, 404n17, 425,

672n20. *See also* Underdeter-
mination, and induction

Unification, 803n15, 994, 996, 1147,
1227, 1253–54, 1260

as criterion for theory assessment,
41, 129–30, 135, 229, 741,
1136, 1187

*Universals, 859–62, 885, 887–90,
896, 898–99, 900n4

Uranus, 28, 31, 71, 270, 304, 314,
461, 463, 577

*Value neutrality, 85, 170, 173, 197,
204, 213–14, 240, 243, 247

Values, 83–85, 170, 211. *See also*
Criteria of theory choice, as
values; Paradigms, normative
functions of; Problems, scien-
tific, standards for the solution
of; Value neutrality

cognitive (epistemic), 84, 85, 140–
56, 157, 160–61, 162, 163,
167n2, 169n38, 186, 211–13,
217, 224–36, 247–50, 391,
401, 640, 668, 1109–10

communal vs. individual, 112–13,
116, 157–59, 168n12, 186,
250

conflict between, 160–61, 182,
185–86, 212, 226

contextual, 85, 173, 186, 197, 204,
212–14, 242–43, 245–50

as ends and means, 129–31

justification of, 128–31, 230–31,
232–33, 236, 250

moral, 211, 251n1

and paradigms, 88, 121–22, 124,
126, 140–44, 168n12, 216,
220, 225–28, 230–32

and problem of induction, 112–14

rationality of. *See* Rationality, of
values

stability and change of, 114–15,
124, 125–29, 140–41, 142–56

Verifiability, 820, 843n2, 1268

criterion of meaning. *See* Meaning,
verifiability criterion of

as demarcation criterion, 27, 29–
30, 31, 64, 72

and intuitionism, 379

Verification, 5–7, 63, 113, 431, 510,
1091, 1194, 1202

theory of meaning. *See* Meaning,
verification theory of

- *Verisimilitude. *See* Truth, approximate
- Vienna Circle. *See* Positivism, logical
- Vitalism, 430n, 768, 993. *See also* Entelechy
- Vulcan, postulation of, 304
- Weak force/interaction, 13, 67, 449, 1159–60, 1165, 1168n13, 1175, 1263
- Wholism. *See* Holism
- Women, 183, 192, 194–95, 196
 - not archetypal humans, 197–98
 - epistemic privilege of, 199, 206, 247–49
 - in science, 183, 244
 - skeletal ideal of, 197–98
- World view(s). *See* Paradigm(s); Research programmes, scientific
- X rays, 86–87, 122–24, 231

The most comprehensive anthology of its kind, *Philosophy of Science: The Central Issues* offers an exceptionally balanced introduction to the main currents in twentieth-century philosophy of science. Forty-nine articles and self-contained excerpts by Thomas S. Kuhn, Karl Popper, Carl G. Hempel, W. V. Quine, and other prominent philosophers survey the foundational questions in the field: What distinguishes science from pseudoscience? Is scientific change a rational process? In what sense is science objective? What are natural laws, and what do they describe? Are scientific theories to be understood as offering a true account of the world? and others.

To help uninitiated readers navigate the sometimes challenging literature in the field, the editors of this collection provide unusually extensive apparatus. Brief introductions to the nine thematic sections sketch out the major issues addressed in the subsequent readings. Fair-minded and thorough commentaries following the readings discuss the selections in depth and draw connections to other pieces in the anthology; they also develop salient arguments, explain unfamiliar theories, and place the readings in a wider philosophical context. A bibliography, a glossary, and two indexes—one for subjects, the other for names—further enhance this volume's value as a reference.

ABOUT THE EDITORS

Martin Curd is associate professor of philosophy at Purdue University. After taking a B.A. in natural sciences at Cambridge University, he studied history and philosophy of science at the University of Pittsburgh, where he received his Ph.D. His articles on a wide range of topics in philosophy of science and epistemology have been published in journals such as the *British Journal for the Philosophy of Science*, *Ratio*, and *The Proceedings of the Philosophy of Science Association* and have been reprinted in several anthologies, including *Readings in the Philosophy of Science*, Second Edition (Prentice Hall, 1989). He is coauthor of *Principles of Reasoning* (St. Martin's Press, 1989) and author of *Argument and Analysis: An Introduction to Philosophy* (West, 1992).

J. A. Cover is associate professor of philosophy at Purdue University. Leaving a research post after completing a B.S. in biochemistry and biophysics at the University of California, Davis, he took a B.A. at Syracuse University, where he later received his M.A. and Ph.D. His primary fields of research are early modern philosophy, metaphysics, and philosophy of science. His work on Leibniz, Spinoza, causation, space and time, and modality has appeared as book chapters and in numerous journals, including *Studies in History and Philosophy of Science*, *Noûs*, *Synthese*, and *Philosophical Studies*. In addition, he is coeditor of *Central Themes in Early Modern Philosophy* (Hackett, 1990) and coauthor of *Theories of Knowledge and Reality: An Introduction to the Problems and Arguments of Philosophy*, Second Edition (McGraw-Hill, 1994).

W • W • NORTON



NEW YORK • LONDON

