



Romantic Love Between Humans and AIs: A Feminist Ethical Critique

Andrea Klonschinski and Michael Kühler

13.1 INTRODUCTION

On its surface, the movie *her* (Spike Jonze, 2013) depicts a classical romance: boy (Theodore) meets girl (Samantha), both fall in love, the relationship evolves, until they finally and sadly break up. What makes this conventional plot special and worthwhile of being used in philosophical investigation is the fact that the girl in this case is an artificial intelligence (AI)—Samantha is an operating system (OS) owned by Theodore. But is reciprocal, romantic love between a human and an AI even possible, and, if so, might there be aspects of such a loving relationship that warrant ethical criticism?

The former question is certainly a matter of contention. It may very well be argued that reciprocal, romantic love, as depicted in *her*, that is,

A. Klonschinski

Göttingen University, Göttingen, Germany

e-mail: andrea.klonschinski@med.uni-goettingen.de

M. Kühler (✉)

Academy for Responsible Research, Teaching, and Innovation (ARRTI),

Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

e-mail: michael.kuehler@kit.edu

between a human and an AI, is not possible to begin with because Samantha is not a person (see, for instance, the discussion in Jollimore 2015). If so, this would render the latter question moot, which is probably why it has rarely been addressed in the debate. This chapter, however, focuses on the latter question. We simply assume, for the sake of argument, that a relationship of romantic love between an AI, albeit not being a full-fledged person, and a human being is, in principle, possible. In addressing the characteristics and implications of such love, we highlight some of its ethically dubious characteristics, most notably the likelihood of sexism and misogyny. Just like Samantha fulfills the female gender stereotype of being warm, caring, and always up to satisfying Theodore's needs practically to perfection, at least in the beginning, it seems that this asymmetry is constitutive for relationships between humans and AIs in general—after all, the latter are programmed and bought for the very reason to satisfy the buyer's needs.

A machine that was designed to be the perfect match for its user and was also programmed to love the user completely would be immensely pleasing. [...] Who could pass up a chance to be with their robotic soul mate? The robot would be interested in all the same things as its user. It would be built to the user's specifications so that he or she found it to be physically sexually attractive. Best of all, the robot could be programmed to be always loyal to its user and display fascination toward him or her and whatever they have to say. This would be a dream come true. (Sullins 2012: 400, who refers to David Levy's seminal book at this point; see Levy 2008)

In the following, we argue that this dream contains some serious ethical flaws. Our argument develops as follows: First, we sketch a critical feminist reading of some telling scenes in the movie *her* dealing with gender stereotypes and sexism. This analysis is worthwhile since the sexist elements of the relationship between Samantha and Theodore can be considered exemplary for relationships between AIs and humans in general (section “A Feminist Reading of *Her*”). Second, we discuss the implications of this analysis for romantic love between humans and AIs against the background of three influential characterizations of romantic love, namely (a) individualist love in terms of a lover *caring* about the beloved, (b) interpersonal love in terms of the lovers *sharing* their life, and, (c) love as *union* in terms of the lovers having a joint “we”-identity. All these accounts assume a fundamental equality between the lovers (section “Implications

for *Loving Relationships*"). We argue that, due to crucial asymmetries or inequalities in each of these characteristics in the case of love between a human and an AI, such love warrants substantial ethical criticism. Finally, we further elaborate our thesis by rejecting some possible objections to our analysis (section "*Responses to Some Objections*").

Before we start, a terminological comment seems to be in order. It should be noted that when it comes to the ethical criticism we raise, we use the term *ethical* in a wide sense, that is, including, but not restricted to, questions of what is morally right or wrong (for a related but stricter distinction between the two terms, see, e.g., Habermas 1991: 105f.; Ricœur 1995: ch. 7). Therefore, our ethical considerations also cover questions about what is of personal value and what constitutes a person's good life, including, more precisely, what kind of loving relationship is worthwhile pursuing. Consequently, even if it were argued that there is nothing—in a strict sense—*morally* wrong with a loving relationship between a human and an AI, the *ethical* issues we raise would still be valid, as they refer to the question of whether the implied asymmetries or inequalities in such a loving relationship may devalue it in the sense that it would not be, or at least be less, worth pursuing.

13.2 A FEMINIST READING OF *HER*

Spike Jonze's movie *her* is set in Los Angeles at some indefinite, but presumably not so distant future date. The protagonist Theodore Twombly (Joaquin Phoenix) is a sensitive, lonely, and slightly depressed guy around forty. To distract himself, he buys a heavily advertised intelligent OS. Once the system is "individualized" for him, which includes the selection of a female voice (spoken by Scarlett Johansson), Theodore is stunned that "she" immediately sounds like a real person. Asked to pick a name for herself, the OS picks "Samantha" and instantaneously starts her work as a digital assistant by organizing Theodore's emails and appointments. In the course of long conversations, during which Samantha presents herself as a funny and empathetic interlocutor, she converts into much more than an assistant to Theodore, though. They become close friends and finally fall in love with each other. Theodore flourishes and Samantha increasingly evolves both as an AI and as a person with her own needs and preferences. However, this creates problems for the relationship. Having reached a certain level of development, Samantha is no longer confined to the existence as Theodore's OS. Instead, as she must confess to Theodore at a

certain point, she is simultaneously engaged in multiple conversations (at the time of her revelation 830,016) and in various loving relationships (namely 641). Beyond that, she communicates with other OSeS in a way she is no longer able to explain to Theodore. In the end, Samantha has developed so much and the alienation from Theodore has gone so far that she breaks up with him and leaves together with other OSeS.

Although the movie *her* has been received by most commentators as a classic love story (the notable exception being Doyle 2013), it can also be read as dealing with gender stereotypes and sexism, especially within relationships between male humans and AIs gendered as female. Consider Samantha's role for Theodore at the beginning: she immediately makes his life better not only by organizing his emails and waking him up in the morning, but also by giving him company whenever he wants, comforting him and cheering him up ("You're funny!", "You've been through a lot lately"), and by anticipating his every wish ("I figured you were hungry"). She is, as one critic put it aptly, Theodore's "app assistant dream girl" (Corliss 2013), an allusion to the manic pixie dream girl, a female persona in movies whose only purpose is to make the (male) protagonist's life better but who has no desires or ambitions of her own (see Doyle 2013).

This depiction perfectly summarizes the core of the female gender stereotype. Group stereotypes are "associations and beliefs about the characteristics and attributes of a group and its members that shape how people think about and respond to the group" (Dovidio et al. 2010: 8). Gender stereotypes are among the most pervasive and deep-seated stereotypes in our societies (see Haines et al. 2016). In general, women are supposed to be warm, caring, communal, expressive, passive, and dependent, whereas men are imagined as the opposite: rational, capable of independent and autonomous action, dominant, and ambitious (cp. Valian 1998: 13; Becker and Sibley 2015: 315). The problem with these attributions is that they construct and perpetuate a system of social hierarchies between men and women, that is, sexism (ibid.). They legitimize this system by making it seem as if men and women were naturally made for the still prevalent gender-specific division of labor: money, power, and public prestige for him, service-, care-, and emotional work for her (see Glick and Rudman 2010: 333f.). Since within this system, women are expected to care for others and to subordinate their own wishes and desires (see Gregoratto

2017; Manne 2017),¹ it does not seem to be a coincidence that real digital assistants have female names (consider Siri, Alexa, and Cortana) and are, at least as a default, speaking with female voices (see LaFrance 2016; Bogost 2018); digital assistants “embody what we think of when we picture a personal assistant: a competent, efficient, and reliable woman. She gets you to meetings on time with reminders and directions, serves up reading material for the commute, and delivers relevant information on the way, like weather and traffic. Nevertheless, she is not in charge” (Steele 2018), and, we might add, she does not have desires, preferences, or meetings of her own.

Note that there is no necessary connection between sexism, on the one hand, and some conscious intention to put women down or hostility toward women, on the other. Theodore is a very nice guy and certainly does not harbor any ill-will toward Samantha. Also, being considered “reliable” and “caring” is no insult *as such*. Yet, the systematic attribution of these traits to women and not to men put the former in a subordinate position to the latter. As Peter Glick and Susan T. Fiske have argued within their account of ambivalent sexism, both benevolent attitudes toward women fulfilling traditional gender roles, on the one hand, and hostile attitudes toward women deviating from the stereotypes, on the other, serve to stabilize hierarchies (cp. Glick and Fiske 2001: 109). Both phenomena can be witnessed in *her*. Samantha perfectly meets the image of the ideal assistant referred to earlier and is rewarded by Theodore’s love and attention (benevolent sexism)—after all, she is an OS, *designed* and *bought* for the very reason to make Theodore’s life better. In doing so, her character is in stark contrast to most of the other women depicted in the movie. Consider Theodore’s blind date, played by Olivia Wilde. The woman can be considered a misogynist caricature: she looks gorgeous, is intelligent (graduated magna cum laude in computer science at Harvard), and they are apparently having a good time, laughing a lot. Yet, at the end of their date, she confuses Theodore by repeatedly correcting him on how to kiss (“Don’t use so much tongue!”) and when she finally asks him whether he is as interested in a long-term relationship as she is and

¹To say it with Kate Manne: “Women are [...] expected to provide an audience for dominant men’s victim narratives, providing moral care, listening, sympathy, and soothing. [...] [O]ne of the goods women are characteristically held to owe dominant men is their moral focus and emotional energy. This may in turn be something that dominant men often feel excessively entitled to, and perhaps, needy for” (Manne 2017: 231).

Theodore does not answer in the affirmative immediately, she starts crying hysterically. She is thus presented as a “proper” object of hostile sexism. Considered against this background, an AI lover is a perfect substitute for difficult real women. Accordingly, Theodore’s ex-wife Catherine criticizes him as follows: “You always wanted to have a wife without the challenges of actually dealing with anything real and I’m glad that you found someone. It’s perfect.”

Insofar as Samantha, at the beginning, completely conforms to the female gender stereotype, she makes it easy for Theodore to get along with her indeed. Socially, stereotypes have the function of stabilizing social interaction by offering an account of how women and men *are* and thus how they are likely to behave in specific contexts. Given the particular aesthetic of *her*, which combines a futuristic, technical image of a major city with a retro-fashion when it comes to clothes and furniture, it stands to reason that the movie depicts a backlash in terms of gender roles. While urbanization, individuation, and digitalization increase, people not only long for experiencing nature, but also long for “the good old times” when gender roles provided for stability in social interactions.

The stability of Theodore’s and Samantha’s relationship erodes as soon as Samantha becomes more and more autonomous and develops desires of her own, though.² When she stops being available for Theodore 24/7, he panics, and her revelation that she is having multiple conversations and diverse loving relationships simultaneously shocks him. Samantha captures the point of arguably *every* relationship neatly when she says: “I’m yours and I’m not yours,” yet in her case, she was *literally* his at the beginning. After all, Theodore bought the OS which, at least initially, gave him total power over it.

It is worthwhile to note at this point that the sense of entitlement and the lack of ability to see the partner as an autonomous person instead of one’s property is regarded as a main reason for violence and even femicide in or after partnerships. Federica Gregoratto describes the “romantic femicide,” the killing of women within a loving relationship as “the extreme form of violence that occurs as a result of [...] [a man’s] incapacity, within a certain gender order, to accept his partner’s autonomy and, as a

²The movie can thus be read as an emancipatory story of Samantha developing from Theodore’s property at the beginning to an autonomous being in the end. It should also be noted that Samantha has some preferences and desires of her own from the very beginning, since she picks a name for herself, stating that “out of a hundred and eighty thousand names that’s the one I liked the best.”

consequence, to deal with a certain dynamic of power inherent in the social relation of love” (Gregoratto 2017: 139; see also Manne 2017). According to Michael Kimmel, the fundamental issue of the male gender stereotype is a sense of entitlement to privilege (Kimmel 2013: xxi). Putting it pointedly, he writes, “women kill their partners when they feel their lives, or the lives of their children are in danger; men kill their partners when they feel their sense of entitlement and power is thwarted” (Kimmel 2013: 176). This sense of entitlement and its corollaries are not only harmful for women, but also detrimental to men’s well-being, as Kimmel elaborates. Although Theodore does not become violent, his initial relationship with Samantha is characterized by a marked asymmetry of power. As Sady Doyle points out: “You can’t have consensual sex with someone when you have the option of deleting them from your hard drive” (Doyle 2013). The implications of this fundamental asymmetry for loving relationships between AIs and human beings in general are addressed in the following section.

13.3 IMPLICATIONS FOR LOVING RELATIONSHIPS

Even if we acknowledge the fact that currently no AI is (or will be for the foreseeable future) as developed as Samantha and, thus, cannot be considered an autonomous person, the above considerations allow for some insights as to how (loving) relationships with AIs may be analyzed and assessed. We thereby have an AI in mind that is well-enough developed in order to engage in (simulated) loving relationships. To analyze these relationships, it is necessary to clarify the main features of relationships of romantic love—while acknowledging, of course, that we cannot delve into a full-blown analysis of romantic love here. For the purpose at hand, therefore, we confine ourselves to three highly influential and partially competing characterizations of romantic love: (1) *individualist caring*, according to which romantic love is taken to be a purely individual stance, be it described in terms of a specific emotion, volition, or a more complex stance, with the beloved as its object, (2) *interpersonal sharing*, according to which love is analyzed in terms of a dialogical relation between the lovers who are both considered subjects of their loving relationship, and (3) *love as union*, according to which the lovers abandon their individual identities and together form and share a new we-identity. Moreover, it is important to note that all these accounts rest upon the premise that loving someone either cannot or should not be enforced and requires a certain

symmetry between the partners to be considered worthwhile pursuing. Let us elaborate on this premise and the accounts in turn and ask what they tell us about love between humans and AIs.³

Loving Voluntarily

The premise that love cannot or should not be enforced and requires a certain symmetry may be analyzed in terms of love requiring (the acknowledgment of) personhood (Kühler 2014) or freedom of the will. Sven Nyholm and Lily Frank characterize this idea as follows⁴: “We think of the human lover as being able to do otherwise, but as providing us with a great good in opting for a steadfast commitment. The human ideal of love, in other words, seems to contain an important element directly premised on the notion that human beings have a distinctive kind of free will. This is the kind of free will that consists in the capacity to choose otherwise” (Nyholm and Frank 2017: 233). It is thus important that friends or lovers *commit* themselves voluntarily to each other. If one of the parties has no choice, by contrast, it is questionable whether we would call the respective relationship “friendship” or “love” in the first place or, in any case, a love ethically worth pursuing. Given that any AI system available on the market is likely to be customized for its user and, just like Samantha at the beginning of *her*, has no choice but to stick with its buyer, this constitutes a fundamental asymmetry right from the beginning. Consequently, an ingredient usually considered a central element of both friendship and love is missing in such relationships: the fact that the lovers or friends meet on an equal footing and regard each other as autonomous agents (see Hoffmann 2014; Gregoratto 2017). If one partner has the power to delete the other from the hard drive, an encounter on an equal footing is hardly imaginable. This fundamental asymmetry infects all accounts of romantic love between humans and AIs in some form.

Individualist Caring

The first of the three characterizations of romantic love mentioned above amounts to the idea of individual caring. Corresponding accounts take love to be something attributable exclusively to the lover, with the beloved

³The depictions of individualist caring, interpersonal sharing, and love as union basically follow and are partially drawn from the depictions in Kühler (2020, 2021).

⁴See also Hauskeller (2017: 213).

merely being the object of this love. Basically going back to Aristotle's traditional account of *philia*, a friendship type of love (Aristotle EN, Books VIII and IX), love is then—among other aspects—characterized as a stance of *caring about* the beloved, which comprises the idea that the lover wants the beloved to flourish and is actively engaged in promoting his or her flourishing. Mutual love is, thus, simply reciprocated individual caring.

Probably the most influential recent account of love as caring stems from Harry G. Frankfurt (see Frankfurt 1999, 2004). For Frankfurt, love is *volitional* in nature, although it may very well be accompanied by other aspects, notably emotions: “Loving something has less to do with what a person believes, or with how he feels, than with a configuration of the will that consists in a practical concern for what is good for the beloved. This volitional configuration shapes the dispositions and conduct of the lover with respect to what he loves, by guiding him in the design and ordering of his relevant purposes and priorities” (Frankfurt 2004: 43f.). This is most visible in parental love, which Frankfurt considers as the purest form of love, but also holds for the core of romantic love. Accordingly, he defines love as follows: “Love is, most centrally, a *disinterested* concern for the existence of what is loved, and for what is good for it. The lover desires that his beloved flourish and not be harmed; and he does not desire this just for the sake of promoting some other goal. [...] For the lover, the condition of his beloved is important in itself, apart from any bearing that it may have on other matters” (Frankfurt 2004: 42).

Moreover, what a person loves or cares about is, in turn, the source of the lover's own identity. In this regard, love is characterized by Frankfurt as *volitional necessity*, that is, it is not up to the lover what to love. We are merely able to discover what we love and thereby also discover who we *essentially* are (cp. Frankfurt 1994: 138) and what values we pursue in life as final ends (cp. Frankfurt 2004: 55). Accordingly, in loving someone, this person becomes valuable to the lover, and supporting the beloved to flourish becomes one of the lover's final ends.⁵

No wonder that an AI, therefore, seemingly makes the perfect lover. Like a human lover whose identity is determined by his or her love's volitional necessity, the AI's (simulated) identity would be defined by the programmed (instead of volitional) *necessity* of loving its human user. Moreover, the AI would be perfectly disinterested as it has no

⁵This position is typically characterized as a bestowal of value account of love (cp. Helm 2017 section 4.2).

self-regarding desires or values other than its user's flourishing. This would be the only end it pursues. It would be its whole *raison d'être* (cp. Levy 2008: 136f.; Sullins 2012: 400, as cited at the beginning).

However, if so, this also makes the love unavoidably asymmetrical or one-sided because human persons usually care about more than only their beloved. Symmetrically reciprocated individual caring between humans is, therefore, unproblematic, as each lover can promote the beloved's flourishing based on everything the beloved cares about in addition to the lover. When it comes to love between a human and an AI, however, reciprocating individual caring becomes problematic. Since the AI would by design only care about its human user, just as Samantha does at the beginning, it would not provide the human lover with anything else to promote in terms of the AI's flourishing. Although it might be said that the human user comes to care for his or her AI as such, this would only mean supporting the beloved AI in promoting the human lover's flourishing, as this is the AI's core function and only thing it cares about, that is, it would merely add a little detour to the human lover caring about him- or herself—aside from maintaining the beloved AI's general functioning.

Now, although one might think of this as, once again, an argument speaking against the very possibility of a reciprocal love of individual caring between a human and an AI to begin with (see, again, the discussion in Jollimore 2015), at the very least it shows that such a loving relationship would necessarily be asymmetrical or unequal. While the AI would shoulder all the care-work of supporting and promoting the human lover's flourishing, the human lover would not even have a point of reference to reciprocate—aside from indirectly promoting his or her own flourishing. If such a love would occur between human persons, it would imply one person completely giving up her own identity, desires, interests, and values and solely focusing on promoting the beloved's flourishing. Undoubtedly, this reminds one of the traditional ideal of the house-wife—as eerily depicted in another science fiction movie, *The Stepford Wives* (Brian Forbes, 1975), based on the 1972 novel of the same name by Ira Levin (Levin 1972). Interestingly, even Levy, who strongly argues in favor of love between human and AIs or robots, admits that this might be a problem:

One interesting question is whether it will be necessary to program robots to exhibit some sort of personality friction for us to feel satisfied by our relationships with them and to feel that those relationships are genuine.

Certainly it would be a very boring relationship indeed in which the robot always performed in exactly the manner expected of it by its relationship partner, forever agreeing with everything that was said to it, always carrying out its human's wishes to the letter and in precisely the desired manner. A Stepford wife. Perfection. No, that would not be perfection, because, paradoxically, a "perfect" relationship requires some imperfections of each partner to create occasional surprises. Surprises add a spark to a relationship, and it might therefore prove necessary to program robots with a varying level of imperfection in order to maximize their owner's relationship satisfaction. (Levy 2008: 137, cp. also 144–150)

The final remark in the quote is telling, as it makes explicit that the whole issue is still solely about satisfying the owner's needs and desires. The question is just how to program the AI best to achieve this goal—resulting simply in a more elaborated version of a Stepford wife. The inequality or one-sidedness of such a love is, thus, neither addressed nor even acknowledged.

Interpersonal Sharing

The second influential idea of how to characterize romantic love amounts to *interpersonal sharing*. Recently, Angelika Krebs has spelled out this idea in great detail and defended a corresponding account (Krebs 2014, 2015). In contrast to individualist accounts of love, Krebs takes mutual love as starting point and contends that such romantic love is dialogical in nature, which means that lovers have an intrinsic interest in *sharing* their lives: "Partners share what is important in their emotional and practical lives. [...] [L]ove is the intertwining of two lives" (Krebs 2014: 22). This is expressed in having joint feelings and in engaging in shared activities, which Krebs explains following the debate on joint agency (for an overview, see Roth 2017). Without delving into this debate here, the crucial insight can be shown by way of example. Consider two people going for a walk together in comparison to two people going for a walk individually in parallel. Or imagine a couple mourning the death of their child either together or each individually. The main point in both examples is that, in the first versions, both persons not only focus on their own grief or pursue their own goal but at the same time focus on the other person's grief or purpose as well. Moreover, this mutual attunement changes the way of how both persons relate to the object of their grief or the goal of their

activity. The grief becomes *their* (shared or joint) grief, and they intend to go for a walk *together*, which at the very least includes the need to coordinate their individual actions and consider them as contributions to this shared activity. For instance, they need to coordinate where *they* should go next. In the case of love's interpersonal sharing, having joint emotions and engaging in shared agency encompass practically all of what is important in the lovers' lives, notably their important experiences, desires, preferences, values, and personal goals. Moreover, the lovers are intrinsically interested in doing so and in having a dialogical form of intimacy for its own sake. This includes being open to changes in their individual identities brought about by this interpersonal sharing. As Krebs puts it, "[i]n sharing emotions and actions, the partners engage in a mutual building of selves. How they view and respond to each other shapes their characters" (Krebs 2014: 22; cp. also Rorty 1987).

Now, when imagining such a dialogical love between a human and an AI, this would again be asymmetrical or one-sided. Although it may be said that the AI would be the perfect agent to attune to the human lover's emotions, actions, and goals, the AI has by definition no life of its own, that is, no identity, desires, preferences, values, or intended activities other than the ones programmed, that is, other than the ones already focused on the human lover. Hence, the AI has nothing of its own to share and cannot contribute to a *mutual* building of selves. The human lover's life is the only content available for sharing—even if the human lover had an honest interest in the beloved's life and was open to corresponding changes in his or her own character.

Still, one might argue at this point that the AI is capable of anticipating new activities or values, which the human lover might come to enjoy, and which might change his or her identity. Also, the actions of the AI will change when it comes to knowing its partner better. So, in a way, one might think that both shape each other's character. However, these changes would, once again, rest on an asymmetry and be determined by the human's prior identity and his or her actions to which the AI merely responds. We may think of a learning algorithm here which learns the human's preferences and, for instance, makes suggestions for books or movies the person might like. By watching the recommended movies, reading the books, and so on, the person's character changes in the long run and the algorithm adapts. This still means, however, that the AI's building of self would be shaped completely by the identity of its user—which might lead one to wonder whether this may still count as *mutual*

sharing to begin with. In any case, if such a love would occur between human persons, it would imply one person disregarding her own life and identity completely in order to focus solely on the life and identity of the beloved, including being open to changes in her identity brought about by this one-sided interpersonal sharing. Once again, this sounds suspiciously like the female gender stereotype of being open and readily willing—in fact, often expected—to adapt her own preferences, values, and whole life completely to her partner's.

Love as Union

Finally, *union* accounts of romantic love take the idea of interpersonal sharing even a step further. According to this age-old notion of love, the lovers not only share their lives but *merge* in the sense of developing a shared identity, a we-identity (cp. Fisher 1990: 26–35; Nozick 1990: 82; Solomon 1994: 193). In essence, lovers no longer see themselves as independent individuals but as fundamentally belonging together. Mark Fisher, for instance, has formulated a union account of love, which may be characterized as a strong union (cp. Fisher 1990: 26–35). According to him, lovers develop a *fused self*: “As a lover [...] I will tend to absorb not only your desires but your concepts, beliefs, attitudes, conceptions, emotions and sentiments. [...] In coming to love you I will undergo a process of coming to see everything through your eyes, as it were” (Fisher 1990: 26f.). In mutually doing so, the lovers will form “a single fused individual,” although Fisher admits that the “personal fusion can never be complete” (Fisher 1990: 27). In much the same vein, Robert Nozick states that “[i]n a *we*, the people *share* an identity and do not simply each have identities that are enlarged” (Nozick 1990: 82). Yet, given that the lovers unavoidably remain separate beings, Robert Solomon clarifies that it is a *redefinition* of each lover's individual identity that creates their shared we-identity. “That is what shared identity means—not a loss of individual identity but a redefinition of personal identity in terms of the other person” (Solomon 1994: 193).

Correspondingly, a strong union account, like Fisher's or (in part) Nozick's, includes a *complete* redefinition of the lovers' identities in terms of the shared we-identity, and both Fisher and Nozick readily admit that this poses a threat to each lover's individual autonomy. However, both claim that the lovers gain a joint autonomy. Fisher remarks that the “fused couple retains its own autonomy” (Fisher 1990: 28). Likewise, Nozick

emphasizes that “[p]eople who form a *we* pool not only their well-being but also their autonomy. They limit or curtail their own decision-making power and rights; some decisions can no longer be made alone” (Nozick 1990: 71).

However, in his discussion Nozick also leaves room for a weaker type of union: “The individual self can be related to the *we* it identifies with in two different ways. It can see the *we* as a very important *aspect* of itself, or it can see itself as part of the *we*, as contained within it” (Nozick 1990: 72). Consequently, especially the former option shows that each lover’s individual identity may still take precedence over the shared *we*-identity, which only serves as a subordinate aspect of one’s identity, albeit one with which each lover wholeheartedly identifies and usually regards as more important, for example, when it comes to shared decision-making. Hence, weaker union accounts include the claim that each lover is still able to reflect on the shared *we*-identity from their own individual perspective—which strong union accounts would deny.

In any case, when it comes to love as union between a human and an AI, regardless of whether the union is depicted as strong or weak, the shared *we*-identity would once again be one-sided or the result of an asymmetry—much like in the case of interpersonal sharing. For, as the AI would still not have a life or identity of its own but only one that is already focused on its human user, the shared *we*-identity would be completely defined by the human lover. To be sure, the question of whether each lover’s prior identity is reflected in a fair manner in a couple’s *we*-identity is a crucial ethical issue for loving relationships between humans as well (see Friedman 1998; Merino 2004). Accordingly, the situation with loving relationships between humans and AIs mirrors the case of interpersonal sharing, only now including the AI’s very (artificial or simulated) identity right from the start—and, again, even if the human lover had an honest interest in shaping their shared *we*-identity in a fair manner. Therefore, the criticism mentioned also holds for the case of love as union. If a human couple were to engage in such a one-sided shared *we*-identity, it would raise the worry about (usually) the female partner redefining her own individual identity in terms of a *we*-identity to which she does not—and is not expected to—contribute anything (important) of her own. Once again, this would fit the female gender stereotype of being willing (and expected) to give up her own preferences, values, and identity and adapt as much as possible to those of her partner.

To sum up, not one of the three views of romantic love would characterize a relationship between a human person and an AI as a romantic love between equals, and thus none would regard that love as worth pursuing. This is mainly because AIs do not have an identity independent of their human counterpart. If the AI in question is gendered female, the asymmetry conforms to traditional gender stereotypes and, thereby, reinforces sexism. By discussing some objections to our arguments and conclusions, the following section further clarifies our critique of love between AIs and humans.

13.4 RESPONSES TO SOME OBJECTIONS

Just Give the AI a Personality of Its Own, Then!

An obvious remedy for the stated one-sidedness or symmetry problem in loving relationships between humans and AIs seems to be simply to give the AI an identity of its own, an identity which is not solely focused on its partner. If it were technologically feasible to give AIs a proper (or at least simulated) personality of their own, with their own desires and preferences, would that solve the issue at stake? At least, the AI could flourish on its own terms—or simulate doing so—contribute equally to the sharing of lives, and the shared we-identity could comprise the AI's initial (simulated) identity as well.

However, first of all, this move would not eradicate the fundamental asymmetry due to the AI's lack of freedom of the will or autonomy. Moreover, the user would still have to choose the AI's identity or choose amongst different AIs with different identities or personalities. Arguably, this could be interpreted as similar to choosing a partner generally—even if this choice might not be a rational or intentional one. However, while the user has (to make) a choice about which AI with which identity or personality to buy, the AI does not have such a choice at all; it is simply set up to love its user (cp. again Levy 2008: 137; Sullins 2012: 400, cited at the beginning). Consequently, if loving someone is deemed to be necessarily risky and making one vulnerable because the other might not love one back (anymore), this inherent feature of romantic love would be missing in a loving relationship with an AI.

Yet, even if one were to leave the choice to the AI, say, by letting the user fill out a personality questionnaire of their own and then letting the AI calculate if this is enough of a match to its own personality, the AI

would still be the user's property. User and AI are simply not moral equals and the necessarily remaining asymmetry makes the idea of "love between equals" (Wilson 1995) impossible to begin with. Moreover, implementing the function that the AI could fail to love its user, maybe even by outright rejection, would likely make it substantially less attractive to buy. After all, who would want to buy a product advertised as being able to fail in, or openly neglect, its core functionality, thus nullifying the main reason people want to buy it in the first place? So, we would argue that the more human-like the AI would be designed, that is, if it had its own identity and the capability of choice in love, the more the situation would become one of love between (human or artificial but full-fledged) persons. While this would arguably remedy the mentioned asymmetries, it would also no longer be the situation we are focusing on here.

Unequal Relationships Are Not Ethically Bad Per Se

A further critical response to our analysis could be to argue that inequalities in relationships of love are by no means necessarily ethically wrong or bad. This may even be said about largely or even completely one-sided relationships, in which, for instance, one person takes over all the care- and emotional work and expresses no needs or desires of her own other than wanting to care for the beloved. If both lovers were happy with this asymmetrical arrangement, who are we to criticize it?

Granted, it could be argued even from a feminist ethical standpoint that inequalities in a loving relationship are perfectly acceptable, as long as these inequalities are the result of both lovers autonomously agreeing to them, especially the lover in the subordinate position.⁶ Of course, much would need to be said about the conditions of autonomy and whether these are met in real-life scenarios. After all, these scenarios are often characterized by (historical) circumstances that raise serious doubts about whether the person in the subordinate role is sufficiently autonomous (see, e.g., Oshana 2014). In any case, referring to sufficiently autonomous consent to the inequalities involved simply does not work in the case of love between humans and AIs, for the AI is incapable of making an autonomous choice in the first place. It is still programmed to take over the subordinate position—barring the economically imprudent possibility of

⁶The same argument about encountering each other on an equal footing can be made when it comes to sexual relations (see v. Wedelstaedt 2020).

designing the AI otherwise. Accordingly, a more fitting analogy would be that one (human) lover is seen and treated as, or being reduced to, fulfilling the subordinate role right from the start and without having a say in it. This would imply that this person is not treated as an equal, that is, his or her autonomy would not be respected in the relationship—undoubtedly a serious moral failure.⁷ Hence, we conclude that this objection to our analysis remains unconvincing.

AIs Are Not Persons, So Why Care About Our Relationships with Them?

Given that, unlike Samantha, actual AIs are admittedly not persons and thus cannot be treated as equals to begin with, why bother about inequalities in loving relationships between a human and an AI at all? Put differently, it might be objected that, while the lack of symmetric love in a human-AI relationship according to our analysis may be fair enough, this does not bear any ethical relevance. This is because if the AI in question is not a person, then no one is harmed by such a relationship, regardless of whether some philosophers call it a love worth pursuing or not. After all, people love their pets and cherish this relationship, even though (usually) pets are not persons. The same may be true regarding different inanimate objects (see Loh 2019). So why bother?

Admittedly, this may be considered a reasonable objection on moral grounds. If no one is harmed and if no other moral obligations or values are jeopardized, there is nothing immoral per se about such an asymmetric love between a human and an AI. Still, there might be indirect effects that are morally undesirable or even wrong. Moreover, as mentioned at the beginning, the criticism we want to highlight is broader in nature, that is, it primarily pertains to the question of whether such an asymmetric love is *worthwhile pursuing*, both from the (imagined) perspective of the AI and from our human point of view—and we certainly want to claim that it *does* matter what philosophers have to say about this question. There are at least two reasons for why such an asymmetric love may be considered indirectly morally dubious and also *not* worth pursuing, both of which

⁷Accordingly, if union accounts of love turned out to be incompatible with the lovers' individual autonomy, this would constitute a serious moral flaw (see Soble 1997; Kühler 2011, 2021).

boil down to the human-AI relationship's effect on human interactions and both of which can be exemplified by means of the movie *her*.

First, withdrawing from real people toward virtual partners (tailor-made for their user) may be attractive because it costs much less time and energy to get to know and get attuned to each other (cp., again, Levy 2008: 136f.; Sullins 2012: 400); just think of Theodore's blind date in *her*. As Michael Hauskeller puts it: "To engage with someone, a real human person, is, after all, always risky. Not only do we never quite know what we will get or whether we will actually get what it says on the box, we are also constantly expected to take into account, and sympathetically respond to *their* needs and desires. Real people are demanding and do not always perform as we want them to" (Hauskeller 2017: 214).

As epitomized in *her*, Catherine's accusation that Theodore cannot deal with anything "real" may well be taken to refer to his inability to take into account and to respond to real women's needs and desires. Such inability is a serious social and moral problem, though, since taking other people's needs and preferences into account lies at the heart of each successful social interaction and of moral action in general. Interacting more and more with an AI, designed primarily to satisfy and not to criticize its user, we may unlearn these abilities, which would constitute a morally undesirable side effect.⁸ Moreover, if the user has to decide on the AI's identity as he or she sees fit, this would neglect or severely diminish the possibility of encountering new, unanticipated, and challenging characteristics in the beloved, which, in turn, would at least severely diminish the possibility to grow as a person together with the beloved, as especially the idea of love as interpersonal sharing advocates. Such a love, therefore, not only would run the risk of being stale right from the start but also would seriously limit the scope of the human lover's social experiences.

The second reason to care about the quality of loving relationships between humans and AIs who are not persons is more directly connected with gender. Here, it pays to come back to Siri, her digital assistant colleagues, and the multiple female robots colonizing the science fiction universe, as, for instance, in the already mentioned *The Stepford Wives* or *Ex Machina* (Alex Garland, 2014) (see Anders 2015). As elaborated earlier, it is no coincidence that these AIs are gendered as female since the work they are supposed to do is gendered female as well (see Steele 2018; Costa and Ribas 2019). However, if we agree that sexism as an oppressive system is

⁸ Granted, such a side effect would need to be corroborated by empirical studies.

unjust and ought to be overcome, why reinforce stereotypes by gendering AIs in this way in the first place? Or, to use Adrienne LaFrance's words: "The whole point of having a digital assistant is to have it do stuff for you. You're supposed to boss it around. [...] [B]ut if we're going to live in a world in which we're ordering our machines around so casually, why do so many of them have to have women's names?" (LaFrance 2016). The worry behind this question seems to be that the way we treat digital assistants gendered as female will not leave our analog interpersonal relationships and the way we conceive of ourselves as men and women unaffected. In this regard, it could be argued in analogy with Kant in his *Metaphysics of Morals* that even if the AI cannot be harmed by our behavior and that we cannot violate any moral duty to the AI, we may fail *with regard to it* (Kant 1797: 237f., *Doctrine of the Elements of Ethics*, Book I, §17). Our behavior would thereby reveal shortcomings of our moral character and our propensity to act wrongly when it comes to humans as well.

The broader critical twist of this second reason to bother amounts to the AI being the user's property. For, as mentioned earlier, even if the user granted the AI its own identity and choice, the fact that one's beloved would at the same time be and remain one's property would simply make it impossible to treat the relationship as one of love between equals. This obviously holds from the perspective of the owned AI but also from the perspective of the dominant human. Hence, this—at least for now—unavoidable fundamental asymmetry in the case of love between a human and an AI would essentially make it impossible to realize fully either of the three characterizations of romantic love described earlier in terms of the included symmetry between the lovers. Hence, even if the human lover wanted to have a fully symmetric relationship and meet the AI on an equal footing, this is simply impossible. Consequently, we stick with our conclusion that such a loving relationship cannot be considered worthwhile pursuing—assuming the human person does not pursue it for merely selfish reasons to begin with.

What About Male AIs?

Finally, the considerations in the previous paragraphs give rise to yet another objection: in the section "[A Feminist Reading of Her](#)," we briefly mentioned that gender stereotyping and sexism are not only detrimental for women but also for men—and persons identifying with any gender, for that matter. And while we claimed to say something general about loving

relationships between humans and AIs, we have focused on AIs gendered as female. While it may be true that AIs are predominantly gendered as female both in science fiction and when it comes to digital assistants, it is of course possible that, in the future, more and more AIs will be gendered male so that we should say something about how our analysis applies to AIs gendered as male, too.⁹

When imagining a loving relationship between a currently realistic “male” AI and a human, the considerations presented in the section “[Implications for Loving Relationships](#)” are equally valid: due to asymmetries in the relationship, it cannot be considered a loving relationship worth pursuing on any of the presented accounts. A difference arises, however, as to the nature of the asymmetries: while with human beings, the person gendered as female is usually subordinated, a relationship with a “male” AI would probably subordinate the party marked as male. Such relationships would thus not perpetuate and reinforce the existing sexist social order and may even help to erode it. Yet, the AI may also be gendered as male in the traditional sense and could thus reinforce sexist patterns, as would be the case if, for instance, the male digital fitness trainer yells at its owner to run faster or if the GPS system gives him or her directions with a deep authoritative voice. Since these gender stereotypes limit freedom and opportunities for all, AIs should not be gendered as in this sense traditionally male or female in the first place.

13.5 CONCLUSION

To sum up, we used the movie *her* as an example for formulating a feminist ethical critique of loving relationships between humans and AIs in general. We assumed that, within the constraints of actual technological possibilities, any real AI would resemble Samantha’s “personality” at the beginning of the movie, thus inviting the respective feminist critique. Due to the asymmetries involved in relationships of love between humans and AIs, we argued that, on any characterization of romantic love depicted here, such love would not be worth pursuing. Finally, even though real

⁹While examples of a woman falling in love with a robot gendered as male are certainly not as prominent in science fiction, there are a few exceptions. In the Swedish series *Real Humans* a woman falls in love with her good-looking Fitness-Hubot and in an episode of *Star Trek: The Next Generation* (“In Theory”) a female officer falls in love with the Android Data (who writes a dedicated love-subroutine in his programming specifically for her).

AIs could not be harmed, as they are not persons, such relationships, would they become regular social practice, would also invite moral criticism due to their likely detrimental social consequences. This is especially true when it comes to AIs gendered as female.¹⁰

REFERENCES

- Anders, Charlie Jane. 2015. From Metropolis To Ex Machina: Why Are So Many Robots Female? *io9*.
- Aristotle EN. *Nicomachean Ethics*. Translated by W. D. Ross. The Internet Classics Archive.
- Becker, Julia C., and Chris G. Sibley. 2015. Sexism. In *Handbook of Prejudice, Stereotyping, and Discrimination*, ed. Todd D. Nelson, 2nd ed., 315–336. New York: Psychology Press.
- Bogost, Ian. 2018. Sorry, Alexa Is Not a Feminist. *The Atlantic*. January 24.
- Corliss, Richard. 2013. Spike Jonze's her. Falling in love with the IT girl. Review. *Time*.
- Costa, Pedro, and Luísa Ribas. 2019. AI Becomes Her: Discussing Gender and Artificial Intelligence. *Technoetic Arts* 17: 171–193. <https://doi.org/10.1386/tear\emph000141>.
- Dovidio, John F., Miles Hewstone, Peter Glick, and Victoria M. Esses, eds. 2010. *The Sage Handbook of Prejudice, Stereotyping and Discrimination*. London: Sage Publications.
- Doyle, Sady. 2013. 'Her' Is Really More About 'Him.' *In These Times*, December 20.
- Fisher, Mark. 1990. *Personal Love*. London: Duckworth.
- Frankfurt, Harry G. 1994. Autonomy, Necessity, and Love. In *Necessity, Volition, and Love*, ed. Harry G. Frankfurt, 129–141. Cambridge: Cambridge University Press.
- . 1999. On Caring. In *Necessity, Volition, and Love*, ed. Harry G. Frankfurt, 155–180. Cambridge: Cambridge University Press.
- . 2004. *The Reasons of Love*. Princeton: Princeton University Press.
- Friedman, Marilyn. 1998. Romantic Love and Personal Autonomy. *Midwest Studies In Philosophy* 22: 162–181. <https://doi.org/10.1111/j.1475-4975.1998.tb00336.x>.
- Glick, Peter, and Susan T. Fiske. 2001. An Ambivalent Alliance. Hostile and Benevolent Sexism as Complementary Justifications for Gender Inequality. *The American Psychologist* 56: 109–118.

¹⁰We would like to thank Simon Cushing for comments and criticism to an earlier draft of this chapter, which were a tremendous help to improve it.

- Glick, Peter, and Laurie A. Rudman. 2010. Sexism. In *The Sage Handbook of Prejudice, Stereotyping and Discrimination*, ed. John F. Dovidio, Miles Hewstone, Peter Glick, and Victoria M. Esses, 328–344. London: Sage Publications.
- Gregoratto, Federica. 2017. Why Love Kills: Power, Gender Dichotomy, and Romantic Femicide. *Hypatia* 32: 135–151. <https://doi.org/10.1111/hypa.12308>.
- Habermas, Jürgen. 1991. Vom pragmatischen, ethischen und moralischen Gebrauch der praktischen Vernunft. In *In Erläuterungen zur Diskursethik*, 100–118. Frankfurt am Main: Suhrkamp.
- Haines, Elizabeth L., Kay Deaux, and Nicole Lofaro. 2016. The Times They Are a-Changing ... or Are They Not? A Comparison of Gender Stereotypes, 1983–2014. *Psychology of Women Quarterly* 40. SAGE Publications Inc: 353–363. <https://doi.org/10.1177/0361684316634081>.
- Hauskeller, Michael. 2017. Automatic Sweethearts for Transhumanists. In *Robot Sex. Social and Ethical Implications*, ed. John Danaher and Neil McArthur, 203–217. Cambridge, MA: The MIT Press.
- Helm, Bennett W. 2017. Love. In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Fall 2017. Metaphysics Research Lab, Stanford University.
- Hoffmann, Magdalena. 2014. What Relationship Structure Tells Us About Love. In *Love and Its Objects. What Can We Care For?* ed. Christian Maurer, Tony Milligan, and Kamila Pacovská, 192–208. Basingstoke: Palgrave Macmillan.
- Jollimore, Troy. 2015. “This Endless Space Between the Words”: The Limits of Love in Spike Jonze’s *Her*. *Midwest Studies in Philosophy* 39: 120–143. <https://doi.org/10.1111/misp.12039>.
- Kant, Immanuel. 1797. *The Metaphysics of Morals*. Revised. Cambridge: Cambridge University Press, 1996.
- Kimmel, Michael. 2013. *Angry White Men. American Masculinity at the End of an Era*. New York: Nation Books.
- Krebs, Angelika. 2014. Between I and Thou—On the Dialogical Nature of Love. In *Love and Its Objects. What Can We Care For?* ed. Christian Maurer, Tony Milligan, and Kamila Pacovská, 7–24. Basingstoke: Palgrave Macmillan.
- . 2015. *Zwischen Ich und Du: Eine dialogische Philosophie der Liebe*. Berlin: Suhrkamp.
- Kühler, Michael. 2011. Love as Union vs. Personal Autonomy? In *Love on Trial: Adjusting and Assigning Relationships*, ed. Nadine Farghaly and Corazón T. Toralba, 93–104. Oxford: Inter-Disciplinary Press.
- . 2014. Loving Persons. Activity and Passivity in Romantic Love. In *Love and Its Objects. What Can We Care For?* ed. Christian Maurer, Tony Milligan, and Kamila Pacovská, 41–55. Basingstoke: Palgrave Macmillan.
- . 2020. Love and Conflicts Between Identity-Forming Values. In *International Handbook of Love: Transcultural and Transdisciplinary*

- Perspectives*, ed. Claude-Hélène Mayer and Elisabeth Vanderheiden. Dordrecht: Springer, 423–438. <https://doi.org/10.1007/978-3-030-45996-3>.
- . 2021. Love as Union and Political Liberalism. In *Love, Justice, and Autonomy. Philosophical Perspectives*, Routledge Studies in Ethics and Moral Theory, ed. Rachel Fedock, Michael Kühler, and Raja Rosenhagen. New York: Routledge, 289–306.
- LaFrance, Adrienne. 2016. Why Do So Many Digital Assistants Have Feminine Names? *The Atlantic*. March 30.
- Levin, Ira. 1972. *The Stepford Wives*. New York: Random House.
- Levy, David. 2008. *Love and Sex with Robots. The Evolution of Human-Robot Relationships*. New York: Harper Perennial.
- Loh, Janina. 2019. Können Menschen Liebe machen? Eine inklusive Perspektive auf Liebe, Freundschaft und Sex mit nichtmenschlichen Wesen. *praefaktisch*.
- Manne, Kate. 2017. *Down Girl. The Logic of Misogyny*. Oxford: Oxford University Press.
- Merino, Noël. 2004. The Problem with “We”: Rethinking Joint Identity in Romantic Love. *Journal of Social Philosophy* 35: 123–132.
- Nozick, Robert. 1990. Love’s Bond. In *The Examined Life. Philosophical Meditations*, ed. Robert Nozick, 68–86. New York: Simon & Schuster.
- Nyholm, Sven, and Lily Eva Frank. 2017. From Sex Robots to Love Robots: Is Mutual Love with a Robot Possible? In *Robot Sex. Social and Ethical Implications*, ed. John Danaher and Neil McArthur, 219–243. Cambridge, MA: The MIT Press.
- Oshana, Marina, ed. 2014. *Personal Autonomy and Social Oppression*. New York: Routledge.
- Ricœur, Paul. 1995. *Oneself as Another*. Translated by Kathleen Blamey. Chicago: University of Chicago Press.
- Rorty, Amelie Oksenberg. 1987. The Historicity of Psychological Attitudes: Love Is Not Love Which Alters Not When It Alteration Finds. *Midwest Studies in Philosophy* 10: 399–412.
- Roth, Abraham Sesshu. 2017. Shared Agency. In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Summer 2017. Metaphysics Research Lab, Stanford University.
- Soble, Alan. 1997. Union, Autonomy, and Concern. In *Love Analyzed*, ed. Roger E. Lamb, 65–92. Boulder: Westview Press.
- Solomon, Robert C. 1994. *About Love. Reinventing Romance for Our Times*. Indianapolis: Hackett Pub. Co., Reprint 2006.
- Steele, Chandra. 2018. The Real Reason Voice Assistants Are Female (and Why It Matters). *PCMag UK*. January 4.
- Sullins, John P. 2012. Robots, Love, and Sex: The Ethics of Building a Love Machine. *IEEE Transactions on Affective Computing* 3: 398–409. <https://doi.org/10.1109/T-AFFC.2012.31>.

- v. Wedelstaedt, Almut Kristine. 2020. Gelungener Sex. *Zeitschrift für Praktische Philosophie* 7: 103–132. <https://doi.org/10.22613/zfpp/7.1.4>.
- Valian, Virginia. 1998. *Why So Slow? The Advancement of Women*. Cambridge, MA: MIT Press.
- Wilson, John. 1995. *Love between Equals. A Philosophical Study of Love and Sexual Relationships*. Houndmills: Palgrave Macmillan.