AI & SOCIETY (2021) 36:637–649
https://doi.org/10.1007/s00146-021-01173-5

OPEN FORUM

Could you hate a robot? And does it matter if you could?

Helen Ryland[1,2]

Received: 7 July 2020 / Accepted: 10 February 2021 / Published online: 15 March 2021
© The Author(s) 2021

Abstract
This article defends two claims. First, humans could be in relationships characterised by hate with some robots. Second, it matters that humans could hate robots, as this hate could wrong the robots (by leaving them at risk of mistreatment, exploitation, etc.). In defending this second claim, I will thus be accepting that morally considerable robots either currently exist, or will exist in the near future, and so it can matter (morally speaking) how we treat these robots. The arguments presented in this article make an important original contribution to the robo-philosophy literature, and particularly the literature on human–robot relationships (which typically only consider *positive* relationship types, e.g., love, friendship, etc.). Additionally, as explained at the end of the article, my discussions of robot hate could also have notable consequences for the emerging robot rights movement. Specifically, I argue that understanding human–robot relationships characterised by hate could actually *help* theorists argue *for* the rights of robots.

Keywords Human–robot relationships · Hate · Robot rights · Discrimination

1 Introduction

This article argues that humans could hate some robots, and that it matters that humans could hate some robots.[1] To defend this argument, the article proceeds as follows. Section 2 outlines why we should care about the arguments defended in this article. I begin by conceding that my thesis is only morally interesting if robots are morally considerable, as only then would it morally matter how we respond to them. I argue that morally considerable robots are *not* a distant possibility; we can make sensible and important moral claims about at least some robots. I then argue that a particularly pressing moral question concerns what relationships we could and should have with these robots. I explain that the existing literature on human–robot relationships focuses only on positive relationship types (e.g., love, friendship, etc.). By considering relationships characterised by hate, this article provides a novel, interesting, and timely addition to discussions of human–robot relationships. With the above in mind, I end Sect. 2 by concluding that we should care about my argument because it makes a significant original contribution to the robo-philosophy literature, and has morally important implications.

Sections 3, 4, 5, 6, 7, 8 then present and defend my central thesis. Section 3 begins by outlining two senses of 'hate'—an everyday sense, where I can hate objects, events, etc., and a more philosophical sense, in which I can be in a relationship characterised by hate. It is only the latter type of hateful relationships that I consider in this article. I explain how relationships characterised by hate are the polar opposite of loving relationships. I then outline three conditions that must be met for x to be in a relationship characterised by hate with y. First, x must desire that things go badly for y. Second, x must view y as being

✉ Helen Ryland
hryland.philosophy@gmail.com

[1] University of Birmingham, Birmingham, UK

[2] The Open University, Milton Keynes, UK

[1] Tasioulas (2019: 51), drawing on a recent UNESCO report, defines a robot as an artificial being that is mobile and interactive, and which can communicate and display some form of autonomy. I will accept this general definition. Throughout this article, I will refer to different types of robot [humanoid sex robots (Sect. 5); dog-like robots (Sect. 6) etc.]. I will argue that humans could hate all of these different types of robot, though some of the justifications for hatred [e.g., that the presence of robots could lead to the dangerous objectification and mistreatment of vulnerable humans (Sect. 5)] are more likely to be directed towards humanoid robots. It is also worth noting that, whilst I consider whether humans could hate robots, I will not consider the converse question of whether a robot could hate a human (in the ways outlined in Sects. 3, 4, 5, 6, 7).

inherently hateworthy. Third, x must maintain their hate for y through either direct or indirect interactions.[2]

Sections 4, 5, 6, 7 argue that all three of the above conditions can be met in human–robot relationships. Namely, humans can desire that things go badly for robots. Humans can view robots as being inherently hateworthy. And humans can maintain their hate for robots through either direct or indirect interactions. Because all three conditions can be met, I argue that humans can (or at least could) have relationships with robots that are characterised by hate.

Section 8 argues that it matters that humans could hate robots in the ways outlined above. This is because human hatred towards robots leaves morally considerable robots at risk of being mistreated (e.g., by being excluded, put in danger, etc.). The section concludes by considering how discussions of human–robot hate have important implications for robot rights.

2 Why should we care about robot hate?

The arguments presented in this article depend upon (at least some) robots being morally considerable.[3] When an entity is morally considerable, it makes sense to make moral claims about them (e.g., that they can be wronged; that we have moral obligations to treat them in certain ways, etc.). Humans and nonhuman animals are morally considerable in this way, whilst toasters and armchairs are not. We cannot sensibly claim that we morally wrong a toaster by lying to it, neglecting it, etc.

My opponent could argue that robots are like toasters and armchairs—they are not morally considerable. If this is so, then my arguments (below) are morally uninteresting. Just like it does not matter if I hate my toaster, it does not matter if I hate a robot (or even could hate a robot).

In response to this initial objection, this section will consider two ways in which we could understand the moral considerability of robots. First, and in line with the above objection, robots are not, and likely never will be, morally considerable. Second, robots either currently are, or will soon be, morally considerable. I will outline evidence which suggests that we ought to favour the second position. As such, I will argue that robots either are, or soon will be, morally considerable, and so discussions of robot hate are relevant, timely, and morally interesting.

Position one: robots are not morally considerable

As Coeckelbergh (2018: 146) emphasises, in discussions of the moral considerability of robots, "…the "default" or "common sense" position denies that machines can ever have moral standing".[4] On this view, robots simply are not the right sort of entities to be morally considerable. For defenders of this view, this is because robots do not meet any of the criteria for having moral standing: "sentience, consciousness, having mental states, having the ability to suffer, and so on" (Coeckelbergh 2018: 146). This 'common sense' view is discussed by, amongst others, Frank and Nyholm (2017: 316–317), Gunkel (2018: 89–91), Sparrow (2002: 313), Sullins (2011), and Torrance (2008).

A related, but slightly weaker, view can be seen in claims that, whilst it is not impossible for robots to have a moral standing, morally considerable robots are only a very distant possibility. Such a view is discussed (and ultimately rejected) by Danaher (2019a: 'Robots can be our Aristotelian friends'). On this view, because the possibility of morally considerable robots is so remote, any claims about how we ought to treat and react to robots can be dismissed as unnecessary, irrelevant, and uninteresting.

For defenders of the above views, my arguments below (Sects. 3, 4, 5, 6, 7, 8) will likely seem unwarranted and unintuitive. For them, my claims about robots are no different to claims made about toasters and armchairs—all three sets of claims are morally uninteresting. In response, I concede that my arguments will not be of interest to those who accept position one: that robots are not, and likely will never be, morally considerable. However, as argued below, we do not necessarily need to accept position one. There are convincing reasons to favour position two: that robots

[2] These relationships do not require reciprocity. As Tistelgren (2018: 8–11) emphasises, hate can be one-sided and unreciprocated. In the same way as there can be unrequited love, there can be unrequited hate. I consequently accept that, in a relationship characterised by hate, the human could hate the robot without the robot being able to hate the human.

[3] Throughout I will talk about (some) robots being 'morally considerable', 'having moral standing', and having 'moral status'. There is some debate regarding whether these three terms refer to the same concept, or to different things. For example, Jaworska and Tannenbaum (2018: introduction) state that "there are two ways of understanding moral status, or what others sometimes call 'moral standing' or 'moral considerability'", implying that the terms are interchangeable. Others argue that an object can be morally considerable without also having a moral status. For example, this is sometimes a position taken in relation to nonhuman animals. For a neat summary of various views on nonhuman animals' moral status, see DeGrazia (2002).
Similarly, there is ongoing debate about whether 'moral status' is a threshold concept, or whether it can come in degrees. For a good summary of this debate, see Jaworska and Tannenbaum (2018: sections 3 and 4).
There is not space here to outline and respond to these debates. My own view is that some robots either currently are, or could soon be, morally considerable and have moral standing/moral status in virtue of holding some morally relevant grounding properties (sentience, emotions, etc.). I also accept that moral status can come in degrees. For a full discussion of my arguments in defence of this position, see Ryland (2020).

[4] Coeckelbergh (2018: 149) himself does not accept the default position.

either are already morally considerable, or will be morally considerable in the near future.[5]

Position two: robots either are morally considerable, or will be morally considerable in the near future

As outlined above, position one argued that robots are not morally considerable because they fail to meet the relevant criteria for having a moral standing. These criteria are typically explained in terms of the possession of certain morally relevant properties: "sentience, consciousness, having mental states, having the ability to suffer, and so on" (Coeckelbergh 2018: 146). To argue against position one, we thus need to demonstrate that some robots either currently do have these properties, or at least will have these properties in the near future.[6]

The claim that robots already have some of these properties is admittedly controversial. Nevertheless, there is at least *some* research which claims to show precisely that. For example, in their 2013 paper, Castro-Gonzales, Malfaz, and Salichs discussed how they have developed an autonomous social robot (Maggie) which they claim can implement fear, and can also display fear-reactive behaviour (such as moving away from a 'fearful' stimuli). They argue that "… Maggie is endowed with a decision making system based on drives, motivations, emotions, and self-learning" (139). If this is so, then Maggie would appear to possess an architecture that enables her to display at least some morally relevant properties. Namely, Maggie could be claimed to have relevant mental states (drives, motivations, emotions, and self-learning), or at least robot equivalents of these states.

Because the above claim is so controversial, many who discuss the moral status of robots instead make the weaker claim that there will likely be morally considerable robots in the near future. This weaker claim is well-discussed by Frank and Nyholm (2017), who state that "…we can imagine future robots sophisticated enough to enjoy a certain degree of consciousness" (313).[7] To support this claim, Frank and Nyholm emphasise that many researchers are either actively working to create robotic consciousness (Prabhaker 2017), or are discussing the conditions that would need to be met for a robot to be conscious (Bryson 2012; Dennett 1994). Further evidence of current attempts to create conscious robots can be seen in the work of Reggia et al. (2019). They argue that.

> "…developing neurocognitive control systems for cognitive robots and using them to search for computational correlates of consciousness provides an important approach for advancing our understanding of consciousness, and… *provides a credible and achievable route to ultimately developing a phenomenally conscious machine"* (Reggia et al. 2019:18, my emphasis).

Given that there is ongoing research into developing robots with relevant properties (consciousness, emotions, etc.), and that this research appears to be making headway (see the Maggie example, and Reggia et al.'s claims about the credibility of creating phenomenally conscious robots), I argue that we should accept position two. We should accept that, if there are not already morally considerable robots (see Maggie), then there could be morally considerable robots in the near future (if certain conditions are met). This position is also accepted by, amongst others, Danaher (2019b), Gordon (2018), and Laukyte (2017).[8]

[5] For the purposes of this article, I will be following Gordon's (2018: section 5) suggestion that the 'near-future' refers to events that happen "within the next couple of decades".

[6] This is not the only way in which we could argue against position one. Coeckelbergh (2009, 2010b, 2014, 2018) argues that, to determine whether a robot has a moral status, we should not only look at whether they have relevant properties (sentience, emotions, etc.). Instead, we ought to take a relational approach. On this view, we could ascribe moral status to a robot in virtue of the moral relations and attitudes that we have towards it. From this, Coeckelbergh (2009: 181) argues that "…humans are justified in ascribing *virtual* moral agency and moral responsibility to those non-humans that *appear* similar to themselves—and to the *extent* that they appear so—and in acting according to this belief". On this view, if and when robots become sufficiently like humans (moral agents), we ought to assign them some sort of virtual moral status. Coeckelbergh accepts that robots may pass this criteria "…now or in the future" (2009: 189).
Coeckelbergh's arguments provide support for my claim that robots either currently have a moral status, or will have one in the near future. I have not considered his arguments in the main text (a) for space reasons and (b) because my opponent could object that position one was explained entirely in terms of moral status properties, and so position two also ought to be explained in these terms, for consistency.

[7] Recall that consciousness was one of the morally relevant properties suggested above.

[8] Danaher (2019b) argues that "…robots can have significant moral status if they are *roughly performatively equivalent* to other entities that are commonly agreed to have significant moral status…. Using analogies with entities to whom we already afford significant moral status, it is argued that the performative threshold may be quite low and robots may cross it soon (if not already)." (Section 1). For further discussion of Danaher's view, see footnote 29.
Gordon (2018: 2) argues that "…based on the enormous prospects for future technological developments, I take it for granted that IRs [artificially intelligent robots] will become moral machines in the future…I attempt to show that if IRs are capable of moral reasoning and decision-making on a level that is comparable with the moral agency of human beings, then one must see IRs not only as moral patients, but also as full moral agents with corresponding moral rights…"
Laukyte (2017: 2) argues that "…if an artificial agent can be described as (i) rational and (ii) interactive, then we can ascribe (iii) responsibility and (iv) personhood to it, and consequently we can recognise it as having rights based on those capacities and attributes…." Laukyte accepts that such morally considerable robots are a possibility in the near future: "My own discussion looks out a bit further into the future by anticipating a world in which the technology will have been built that makes fully intelligent artificial agents already a real-

At this point, my opponent may object that, even if robots *could* be morally considerable *in the near future,* we are not justified in making moral claims about them *now.* In other words, it still does not (currently) matter if I could hate a robot; this will only matter in the future when the robot becomes morally considerable. There are two main responses to this objection. First, as mentioned above, there may already be robots who are morally considerable *now,* at least to some extent (see the Maggie example). It matters that we could hate these robots (for the reasons outlined in Sect. 8).

Second, even if robots will only become morally considerable in the near future, this ought not prevent us from making moral statements about robots now. This is nicely expressed by Neely (2014), who argues as follows:

> "The time to start thinking about these [moral] issues is now, before we are quite at the position of having such beings to contend with. If we do not face these questions as a society, we will likely perpetuate injustices on many who, in fact, deserve to be regarded as members of the moral community" (109).

Similar arguments will be presented in Sect. 8. For now though, it will suffice to reiterate that we can make sensible and important moral claims about at least some robots (those that either are or will soon be morally considerable) now.

One of the most important moral questions we can ask about robots is what form human–robot relationships ought to take. Current research has examined whether robots can be (i) lovers, (ii) companions, (iii) friends, (iv) caregivers, (v) nannies, (vi) teachers, (vii) reverends, (viii) colleagues, and (ix) teammates.[9] All of this research is necessary and important as we need to properly clarify and categorise human–robot relationships to determine how robots fit into our moral community (if indeed they do) and how we ought to treat them as a result. For example, if we accept that we can have reciprocal friendships with robots, then this could entail that we have certain beneficent duties towards them (and they to us) (Tistelgren 2018: 6–8).

The remainder of this article aims to contribute to the ongoing discussions of human–robot relationships, and to add further clarity to these debates. As shown above, the existing literature has focused largely on examining *positive* human–robot relationships (e.g., whether humans and robots can be friends and love one another). What is missing is a discussion of more *negative* human–robot relationships, for example, one predicated on human hate. By addressing this omission (Sects. 3, 4, 5, 6, 7, 8), this article makes an original contribution to the robo-philosophy literature, specifically the literature on human–robot relationships.

In sum, this section has outlined why we ought to care about the arguments that will be defended in this article. I have explained that discussions of human–robot relationships have currently not discussed how these relationships might be characterised by human hate. The arguments of Sects. 4, 5, 6, 7 are thus an important addition to the existing literature. Further, I have argued that robots either are or soon will be morally considerable, and so we can make sensible moral claims about how we ought to treat them. The arguments of Sect. 8—which outline why it matters that humans could hate robots—thus have notable moral implications. To make these arguments, the next section will briefly outline what it means for a human to be in a relationship characterised by hate.

3 Hate

I hate garlic, small talk, and rush hour trains. You probably hate many things too. In everyday life, we use this colloquial sense of 'hate' to express a negative reaction to certain objects, people, events, etc. It is obviously possible for humans to hate robots, in this everyday sense of the word. This, however, is not the type of hate that this article will focus on.

Instead, our focus will exclusively be on *relationship*s that are characterised by hate. This is because, as mentioned in Sect. 2, my interest is in the relationships that humans can have with morally considerable robots. By examining relationships characterised by hate, we can begin to consider how these relationships might be *negative,* and what effects this might have on our treatment of robots.[10]

In the existing philosophical literature, relationships characterised by hate are viewed as the polar opposite of relationships characterised by love (Ben-Ze'ev 2018: 323;

Footnote 8 (continued)

ity, and in this scenario I ask how our relation to these agents should be framed" (15).

[9] For (i) see Nyholm and Frank (2017). For (ii) see Coeckelbergh (2010a) and Marti (2010). For (iii) see Danaher (2019a), Mulvey (2018) and Tistelgren (2018). For (iv) see Borenstein and Pearson (2010) and Sorell and Draper (2014). For (v) see Bryson (2010), Kubinyi, Pongrácz, and Miklósi (2010), Sharkey and Sharkey (2010), van denBroek (2010) and Whitby (2010). For (vi) see Sharkey (2015, 2016). For (vii) see Young (2019). For (viii) see Bernstein, Crowley, and Nourbakhsh (2007). For (ix) see Groom and Nass (2007).

[10] One might object that I could also have some sort of relationship with garlic, rush hour trains, etc. This may be true, but there is still an important difference between a relationship with garlic and a relationship characterised by hate with a robot. Only the latter type of relationship involves interactions with a morally considerable being. As explained in Sect. 2, the robot, but not the garlic, *could* be morally considerable and so it could matter (morally speaking) how we treat these robots.

Kauppinen 2015: 1721–1722). Kauppinen (2015) explains this as follows: whereas love is concerned with seeking the best for a loved one, "if I hate someone, I want him or her to do badly, whether or not it is of instrumental benefit for me. I feel bad if the person does well, get easily angry with him or her, and may be delighted if misfortune befalls him or her" (1721). This explanation of what follows when we hate someone seems intuitively correct, and we can suppose that my hateful responses to someone can vary in intensity depending on how much I hate them. For example, suppose that I mildly hate a colleague. I might want things to go slightly badly (but not terribly) for them, and be more easily irritated by them and their successes than I would normally be for other people. Conversely, if someone is my nemesis, then I may want things to go appallingly for them, I might be perpetually infuriated by them, and actively root for (and perhaps orchestrate) them to suffer misfortunes. This degrees-of-hate idea runs parallel to the intuitive idea that there are degrees of love. For instance, I may love my colleagues, my friends, and my family, but love my family the most. From this, it follows that, whilst I may seek the best for all of my loved ones, I may be particularly invested in seeking the best for my family.

Using this initial idea—that hate is the converse of love—the existing literature goes on to suggest three distinguishing features of relationships characterised by hate. First, x (the hater) must desire that things go badly for y (the hated). As explained above, this desire can vary in intensity. At the weakest level, x may desire that y is embarrassed or ridiculed. At the most extreme level, x may desire that y is annihilated. Fischer et al (2018: 311) argue that *all* of these negative desires (from humiliation to annihilation) ought to be understood in terms of x's desire to destroy y. They claim that "…the emotivational goal of hate is not merely to hurt, but to ultimately eliminate or destroy the target, either mentally (humiliating, treasuring feelings of revenge), socially (excluding, ignoring), or physically (killing, torturing) …" (ibid). Fisher et al.'s analysis, however, is overly strong as it seems to presuppose that *every* instance of x hating y will be connected to x wanting to destroy y. In contrast, I will suppose that, whilst in relationships characterised by hate, x will always desire something bad to happen to y, this desire will not always be connected to annihilation or destruction.

The second defining characteristic is that x must judge y to have an inherently 'hateworthy' nature. Fischer et al (2018: 310–311) explain that the appraisals of hated person(s) have two main features. First, the hated person(s) are viewed as being a threat or inconvenience to the hater. They may be viewed as dangerous, immoral, malicious, evil, etc. These perceived character faults (however small or imagined) are viewed as reasons to hate the hated person(s). Second, the hated person(s)' hateworthy nature is judged to be a stable attribute of them—they are *inherently* dangerous, immoral, malicious, evil, etc. In the eyes of the hater, the hated person(s) will *always* be dangerous, evil, etc. Importantly, these negative appraisals of the hated person(s) will typically be accompanied by feelings of powerlessness and lack of control. The hater (x) believes that the hated person (y) is inherently bad (dangerous, evil, etc.); that y will never change; and that they (x) are in danger of being a victim of y's dangerous/immoral/evil plans and actions. As Szanto (2018: 10–20) explains, these appraisals generate an us–them mentality. The hated person(s) are a dangerous 'them', who are inherently different to the safe 'us' (the hater, and all those who share their hatred). Unlike 'them', the 'us' group are judged to be kind, good, moral, etc.[11]

Finally, in relationships characterised by hate, hate is maintained through interaction.[12] As Fischer (2018: 325–326) emphasises, "hate needs to be fed, either by direct or indirect interactions related to the object of hate".[13] Direct interaction is when the hater has to interact with, or be around, the hated person(s). Indirect interaction is when the hater can discuss the hated person(s) with others who also hate them (the 'us' group, above), or when the hater indirectly has contact with the hated person(s), e.g., by seeing their social media profiles. In both cases (direct and indirect), hate is maintained because the hater can sustain and reinforce their negative appraisals of the hated person(s).[14]

In sum, this section has outlined two senses of 'hate': the everyday sense in which I claim to hate objects, events, etc., and a more specific philosophical sense in which I can be in a relationship characterised by hate. I have explained that it is only the second sense of hate that I am interested in. I clarified how, according to the existing literature, there are three conditions that ought to be met for x to be in a relationship characterised by hate with y. First, x must desire that things go badly for y. This is a characteristic *behavioural tendency* of relationships characterised by hate. Second, x must view y as having an inherently hateworthy nature. This

[11] Again, this is the converse of love. In relationships characterised by love, the loved one is viewed as having a stable, inherently praiseworthy nature. They are also viewed as one of the 'us' in an us–them mentality.

[12] A related point is that hate is typically enduring. Szanto (2018, 2) emphasises that "…hatred tends to robustly linger and habitualize even in the face of long-faded harm and healed wounds…". This sort of hatred can be seen in feuds. As Ben-Ze'ev (2018) explains, when hate is enduring in this way, it takes on a certain profundity or depth that gives the hate meaning. For the purposes of this article, I will not consider the enduringness of hate as a separate characteristic. This is because hate can only be enduring if it is maintained, and if it is maintained, then it is enduring.

[13] Similar views are presented in Ben-Ze'ev (2018: 323–324) and Szanto (2018: 3).

[14] Again, this has parallels to love, which is also maintained through direct or indirect interaction.

is a characteristic *appraisal*, seen in relationships characterised by hate. Finally, x's hatred of y must be maintained through direct or indirect interactions. This condition outlines the characteristic *connections* between the hater and the hated. Sections 4, 5, 6, 7 will argue that all three of these conditions could be met in a human's relationship with a robot.

4 Humans could be in a relationship characterised by hate with robots

Section 3 outlined three conditions that must be met for x to be in a relationship characterised by hate with y: (i) x must desire that things go badly for y, (ii) x must view y as having an inherently hateworthy nature, and (iii) x must maintain their hatred for y through direct or indirect interaction. Sections 4, 5, 6, 7 will argue that all three of these conditions could be met in human–robot relationships. To show this, I will examine how humans *currently* respond to robots, and how these current responses meet these three conditions.[15] As humans can and do show hateful responses towards *current* robots, it is conceivable that we could also be in a relationship characterised by hate with *morally considerable* robots. Recall that Sect. 2 suggested that there either already are morally considerable robots (e.g., Maggie), or that morally considerable robots will exist in the near future. Section 8 will argue that it matters that we could hate these morally considerable robots.

5 Humans could desire that things go badly for robots

Section 3 argued that, to be in a relationship characterised by hate, x must desire that things go badly for y. I explained that this desire can range in intensity, from a desire to humiliate y to a desire to annihilate y. This section will focus on the most intense desire: that humans could desire the annihilation or destruction of robots. There are two reasons for this focus. First, most of the existing literature on relationships characterised by hate does discuss the desire for annihilation. This can be seen in the works of Ben-Ze'ev (2018: 323), Fischer (2018: 325), Szanto (2018: 2–9), and Van Doorn (2018: 321). As the desire for annihilation is commonly referenced in existing discussions, it seems like a viable starting point for our robot discussion.[16]

The second reason to focus on the desire for annihilation, rather than less extreme desires (like the desire that y be humiliated or socially excluded), is because the desire for annihilation makes the strongest and most interesting case for potential robot hate. If we can show that humans could desire that robots be annihilated, it seems likely that we would also be able to say that humans could have the less extreme desires—that the robot be humiliated, socially excluded, etc. We could not make the same argument the other way around, i.e., that, because humans could desire that robots be humiliated, they could also desire that they be annihilated. As there is not space to consider every way in which humans could desire that things go badly for robots, it makes sense to focus on the strongest and most extreme claim. With this in mind, let us examine how at least some humans seem to currently desire that robots be annihilated or destroyed.

Since 2015, there has been a well-publicised 'Campaign against sex robots'.[17] In essence, the campaign argues against the development of sex robots on the grounds that the use of such robots perpetuates dangerous attitudes, such as the objectification of women, and the blurring of sex and rape (as sex robots typically do not consent to sexual acts).[18] What is particularly interesting for our purposes is what the campaign suggests we ought to do in reaction to sex robots. In an article on the campaign website, Florence Gildea and Kathleen Richardson (2017) make the following claim:

> "It might be argued that the solution, then, is to encourage the production of sex robots designed to appear male. But to argue for an equality of the lowest common denominator—where everyone relates to all others as an object—is to exacerbate the problem, not provide a solution".

To me, the above implies that no modifications to the production of sex robots would remove the problems with objectification. Consequently, it seems that Gildea and Richardson are implicitly suggesting that the *only* solution would

[15] It is worth noting that I do not mean to imply that these current responses are the *correct* or *fitting* responses to take towards robots.

[16] It might be objected that one can desire the destruction of something (e.g., a building) without hating it. This is true but irrelevant to the argument defended here. Recall that my focus is not on x hating y, but on x being in a relationship characterised by hate with y. As explained in Sect. 3, for this relationship to hold, x must meet three conditions: x must desire that something bad happen to y, x must see y as inherently hateworthy, and x must maintain their hate through either direct or indirect interaction.

Footnote 16 (continued)

[17] There is also a 'Campaign to stop killer robots' which calls for a ban on fully autonomous weaponised robots. I will not discuss this campaign further because it is difficult to separate concerns about (and hatred towards) autonomous robots from ethical concerns about the use of weaponry, calls for disarmament, and discussions of just war theory.

[18] Discussions of the societal and ethical implications of robot sex can also be found in Danaher (2017; 2019c) and Danaher and McArthur (2017).

be to eliminate *all* sex robots by discouraging or banning the production of sex robots.[19] We can reach this conclusion if we follow the logic of the campaign arguments. The campaign begins by making a claim—the use of sex robots is dangerous (due to concerns about objectification, etc.). This perceived negative evaluation (danger) extends to all sex robots (of all gender appearances), and so all sex robots are viewed as dangerous (in some way). Because the danger of sex robots is so extensive, we should remove or eliminate all tokens of the dangerous object (all sex robots). This argument is logical and structurally sound, even if we do not agree with the central claim (that the use of sex robots is dangerous).[20]

As presented above, Gildea and Richardson's arguments work by emphasising a supposed inequality between humans and robots. First, they accept that sex robots are created to fulfil human needs and desires, whilst the same is not true about the human (who does not fulfil the needs and desires of the sex robot). An upshot of this—as suggested above—is that as humans created sex robots for this aim (human fulfilment), they can also destroy sex robots when said human fulfilment has unintended negative consequences (like sexual objectification). Second, Gildea and Richardson implicitly emphasise the inequality in vulnerability between robots and humans.[21] They suggest that although a sex robot is not harmed when a human uses it, the use of sex robots *can* indirectly harm the most vulnerable humans (e.g., women, children) by creating societal issues (like objectification and issues with sexual consent) that disproportionately put them at risk. If one adopts this line of thought, then it is at least plausible to claim that the inclusion of sex robots in human society is dangerous and problematic, and that the best solution is to remove the sex robots (by destroying and/or banning them).

The above has used the 'Campaign against sex robots' to suggest that current attitudes towards sex robots can cause at least some humans to develop 'hateful' desires to eliminate or destroy all sex robots. If this is so, then this suffices to show that at least some humans can develop desires that things go badly for at least some robots (here, in the extreme sense that the robot be annihilated). As shown above, this desire for destruction seems to follow a basic logic. The robot becomes an object of hate (in the sense that one can desire its destruction) if the robot's inclusion in human society is at least widely perceived as a danger or threat that needs to be eliminated (however, elimination is understood). This same logic can arguably extend beyond sex robots. Simply put, humans could desire that any robot be destroyed if said robot is perceived as a danger or threat that needs to be eliminated. The perceived danger of the robot (or of its use by humans) could be understood broadly and include both minor threats (e.g., robots could have 'offensive' glitches, like accidentally swearing in front of children), and major threats (e.g., robots could collect personal data about human users). It does not seem implausible to suppose that this perception of robots, and the subsequent desire to destroy them, could extend to the morally relevant robots discussed in Sect. 2.

6 Humans could view robots as having an inherently hateworthy nature

Section 3 explained that, to be in a relationship characterised by hate, x (the hater) ought to perceive y (the hated) as having an inherently hateworthy nature. One consequence of this aspect of hate is that all members of the hated group (all who are like y) are tarred with the same brush. If there is something inherently wrong with one token object of hate (e.g., one advanced social robot), then there will be something wrong with all similar tokens (all other advanced social robots). In what follows, I will draw on current negative reactions to robots to explain how humans can come to view robots as having this inherently hateworthy nature.

First, there is the uncanny valley effect (Lay 2015). This occurs when humans find human-like robots eerie and disturbing. This is because, whilst the robot looks human in its features, it may not react, behave, or speak in a naturally human way. As Mathur and Reichling (2016) explain, the uncanny valley effect can cause humans to view robots as inherently untrustworthy. This is a generalised reaction. *All* humanoid robots could be viewed as inherently eerie, disturbing, and untrustworthy *simply because they are humanoid robots.* Such general, negative appraisals can be used to ground hate towards humanoid robots if the robots' inherent

[19] Similar calls—to completely ban production—can be seen in the 'Campaign to stop killer robots'. See footnote 17.

[20] One might object that the 'Campaign against sex robots' does not actively hate sex robots. They do not frame their discussions in terms of hate towards sex robots. Nevertheless, I maintain that the solution implicitly proposed by the campaign—to destroy or ban sex robots—meets our conditions for hate (the desire to eliminate a target object because the object is viewed as inherently dangerous or hateworthy). I thus claim that the campaign does show hate towards sex robots, even if the campaigners themselves do not acknowledge or reference this hate. This is consistent with Szanto's (2018: 5) claim that "hatred is also extreme in the sense that it is extremely rarely experienced or acknowledged as such.".

[21] Here it is supposed that humans are vulnerable and robots are not. It is possible to question the idea that robots are invulnerable. For example, Coeckelbergh (2010a) has argued that: "whereas cyberspace and information technologies may well aim at invulnerability (and perhaps immortality, as in transhumanist versions of techno-utopian worlds), they depend on software and hardware that are very vulnerable indeed" (13). On this view, robots have certain robot-specific needs (to power supplies, updates, virus protection, etc.), and can be harmed if these needs are not met. This susceptibility to harm makes the robots vulnerable, in some sense.

eeriness, disturbing-ness, or untrustworthiness is taken to be dangerous or threatening in some way (see 4.1, above).[22]

This view—that at least some types of robots are inherently eerie, disturbing, and untrustworthy—also seems to extend to robots that are not humanoid in appearance (and so are not part of the uncanny valley effect). For example, the dog-like robots developed by Boston Dynamics are often described as 'creepy' or 'terrifying' (DeCosta-Klipa 2019; Titcomb 2016). This suggests that, for at least some people, *all* robots (humanoid or otherwise) have an inherently disturbing or threatening nature. If so, then this could sustain an us–them mentality (Sect. 3, above) whereby robots are a threatening 'them' who ought to be hated because of the danger that they pose to human society. Once again, as this negative view of robots seems to potentially extend to *all* robots, it could apply to the morally relevant robots discussed in Sect. 2.

7 Human hatred towards robots could be maintained through interaction

It is a unique feature of human–robot relationships that a lot of our preconceptions about these relationships have been developed through fiction. Many, but not all, science fiction and fantasy plotlines about robots present a dystopian view whereby advanced robots clash with humans, and are ultimately viewed as a 'dangerous threat'. Examples of this can be seen in *The Terminator* films, *Westwood,* and *Humans*.[23]

Research by the *Leverhulme Centre for the Future of Intelligence* has emphasised the significant negative effects that these preconceptions have on our views about robots. Cave et al. (2019) surveyed 1078 UK citizens to observe how negative preconceptions of A.I. (created via interaction with media) affect perceptions of artificial intelligence (A.I.). They found that 51% of respondents were concerned that the rise of A.I. will lead to the alienation and obsolescence of human beings, with 45% concerned that there will be an A.I. uprising (ibid: 4). As "25% of respondents explained A.I. in terms of robots" (ibid: 5), this implies that a non-negligible number of people view robots as a danger or threat, in virtue of preconceptions developed through dystopian narratives.[24]

As fiction is a key way in which humans understand, and indirectly interact with, robots, dystopian narratives could be a significant factor in the maintenance of human hatred towards robots. This will be particularly true in cases where humans use their fears and negative preconceptions to avoid directly interacting with social robots. This is because if such humans fail to directly interact with social robots, they are unlikely to be exposed to evidence which could contradict the fears generated by dystopian narratives.[25]

Human hate towards robots (in terms of viewing robots as threats that need to be eliminated, above) can also be sustained through direct interaction with robots. This can happen in cases where direct interactions with robots reinforces the belief that robots pose a specific threat (e.g., that they are unsafe, or unpredictable, etc.). For example, consider the current American response to robots entering the workforce. The rise of robot workers is typically linked to the threat that existing human workers will be made redundant and will experience a worse quality of life as a result. News reports on these fears often frame the reports in terms of 'hate' (Condliffe 2019; Matyszczyk 2019). Indeed, a 2017 study by the Pew Research Centre emphasised that, out of 4135 respondents, "85% of Americans are in favour of limiting machines to performing primarily those jobs that are dangerous or unhealthy for humans…" (Smith and Anderson 2017). This suggests that those who will directly interact with advanced, social robots (the human workforce) can come to view social robots as a threat (to their quality of life) *and* see the robots as expendable (only to be used for dangerous jobs). This latter concession (that robots could do the dangerous jobs) could again be taken to show support for (and potentially reinforce) the 'hateful' idea that robots can (and perhaps should) be harmed, destroyed, or otherwise eliminated from mainstream society.

In sum, Sects. 4, 5, 6, 7 have argued that humans can (and perhaps already do) meet the conditions for being in a relationship characterised by hate with robots. I explained how humans can desire the total elimination of certain types of

[22] This links to our discussion of sex robots. Most sex robots are humanoid in appearance, and many developers are working to make sex robots as realistically human as possible. For example, the 'Emma' robot by AI AI-Tech UK is marketed as having synthetic skin that is heated to human body temperature. The company are also reportedly in the process of making Emma breathe via a chest cavity (Miller 2019). Emma is a viable candidate for the uncanny valley effect—she looks realistically human, but is not human. This could support the 'Campaign against sex robots' arguments (above). It is because Emma looks human (but isn't) that actions performed against her (objectification, rape, etc.) can be dangerous and threatening for real human people (women and children, etc.).

[23] *The Terminator* franchise focuses on battles between the nearly-extinct human race and the 'Terminator' killer robots. The humans and robots see one another as threats to survival. *Westworld* concerns a fictional theme park where humans can pay to act out their (usually darkest) fantasies on robotic hosts (e.g., raping them, killing them, etc.). Some of the robotic hosts eventually develop sentience and seek revenge on their human tormentors. Like *Westworld, Humans* concerns a group of complex robots (synths) who gain consciousness, sentience, etc. Some of the synths see human beings as threats that need to be eliminated, and vice versa.

[24] See also The Royal Society's 2018 "Portrayals and perceptions of AI and why they matter".

[25] A similar view was presented in Dr Hatice Gunes' 2018 Hay Festival talk (summarised at Cambridge University 2018).

robots because said robots are viewed as inherently dangerous, eerie and untrustworthy. I suggested that this hate can be maintained through indirect interaction (via dystopian fiction) and direct interaction (e.g., by robots entering the workforce). As humans can already meet the above three conditions in their relationships with existing robots, it is not unreasonable to suppose that humans could also be in relationships characterised by hate with morally considerable robots. As with Sect. 2, I leave it as an open question whether these morally considerable robots already exist (and so we could already be in relationships characterised by hate with them), or whether morally considerable robots will be developed in the near future. The next section (Sect. 8) will outline why it matters that humans could be in relationships characterised by hate with morally considerable robots.

8 It matters that humans could be in relationships characterised by hate with robots

Most of this article has been concerned with demonstrating that humans could be in relationships characterised by hate with robots. I have explained how current responses to robots meet the conditions for being in such a relationship (Sects. 4, 5, 6, 7), and I have suggested that these same responses could be extended to morally considerable robots either now or in the near future (Sects. 2, 4, 5, 6, 7). This section will outline why it *matters* that humans could be in relationships characterised by hate with morally considerable robots.

First, it matters because it shows that humans could feasibly have negative relationships with robots. This is important because the existing literature on human–robot relationships has focused on *positive* relationships, like friendship, love, etc. (Sect. 2). As human–robot interaction becomes more commonplace, it is vital that we have an in-depth understanding of how these relationships work. Our understanding is incomplete if we do not acknowledge the very real possibility that some of our relationships with morally considerable robots will be negative. For example, my arguments on hate (Sects. 3, 4, 5, 6, 7) suggest that we ought to also consider whether robots can be enemies, opponents, competitors, etc. It is only by considering these negative relationships that we can understand the problems, as well as the benefits, of human–robot interaction. For instance, it is likely that, by further examining human–robot relationships involving hate, rivalry, etc., we will get a better understanding of the conflicts that can arise between humans and robots (in terms of resources, opportunities, rights, etc.).[26]

Second, it matters that humans could hate advanced, person-like robots because there is an imbalanced relation between us (humans—the hater) and them (robots—the hated). At present, it is only us that can create advanced, person-like robots, and it is also us who can destroy robots or eliminate them (by ceasing production). Regardless of how advanced, person-like, or functionally similar robots are to us, or how similar they will become to us (Sect. 2), they are currently not *equal* to us in this regard. Robots' continued existence is entirely dependent upon us and our *good will.* As explained in Sects. 3, 4, 5, 6, 7, relationships characterised by hate explicitly involve *bad will* towards robots (via a desire that bad things happen to the robot). Given the enormous power that we wield over robots, it is important that we acknowledge how hate (bad will) could bias or prejudice our perceptions of them. Their existence could depend upon us doing this.[27]

Finally, and in relation to the above, the fact that humans could be in relationships characterised by hate with morally considerable robots has significant applications for the ongoing robot rights discussions. If morally considerable robots are to enter human society and have relationships with us (Sect. 2), then we ought to be clear about how we should treat these robots (and how they should treat us). An important consideration here is whether morally considerable robots could themselves have rights, in the sense of having claims that others have duties to fulfil. For instance, we might consider whether a morally considerable robot has a legal status, a nationality, a right to privacy, etc.[28]

All of the existing research *in favour* of robot rights is predicated on the following claim: *If robots are sufficiently person-like, then they are morally considerable, and ought to have rights.*[29] The above discussions of hate (Sects. 3, 4, 5,

[26] Research into these conflicts could raise interesting questions about the *legitimacy* or *fittingness* of negative reactions towards robots (hate, rivalry, etc.). If there are human–robot conflicts for resources, opportunities, etc., then some of the negative appraisals mentioned earlier—that robots are dangerous to human welfare—could be justified, legitimate concerns. The legitimacy or illegitimacy of hate/rivalry towards robots is, however, a question for another paper.

Footnote 26 (continued)

[27] This could change if robots become able to create other robots.

[28] For a good discussion of the moral and legal status of intelligent robots, see Gordon (2020).

[29] As in Sect. 2, current accounts accept that a robot is sufficiently person-like if it possesses certain important properties (rationality, emotions, etc.). For many theorists, the relevant properties are understood as some morally salient properties that are held by standard adult humans. This can be seen in the following accounts:

"If machines attain a capability of moral reasoning and decision-making that is comparable to the moral agency of human beings, then they should be entitled to the status of full moral (and legal) agents, equipped with full moral standing and related rights" (Gordon 2018: 3).

"I argue that if a machine exhibits behaviour of a type normally regarded as a product of human consciousness (whatever consciousness might be), then we should accept that that machine has

6, 7) suggest that this claim is too quick. This is because, as outlined in Sects. 4, 5, 6, 7, it is when robots *are* sufficiently person-like (in terms of appearance, behaviour, attributes, or skill-set) that humans can view them as inherently hateworthy and (consciously or otherwise) seek their destruction. Consequently, it is the very conditions for being a rights-holder (being sufficiently person-like) that puts robots at risk of human mistreatment (unjust discrimination, exclusion, destruction, etc.—the very things that rights are supposed to protect against). This is important because it is humans who determine the conditions for being a rights-holder (being sufficiently person-like), and *also* humans who can hate robots when they meet these conditions.[30]

To explain this claim further, we can draw on Manne's (2016) critique of humanism. Manne begins by outlining an oft presented view that inhumane conduct (oppression, genocide, rape, etc.) can best be explained in terms of dehumanisation—x is able to mistreat y because they see y as less than human (and akin to a nonhuman animal, an object, etc.). In reaction to this view, Manne suggests that at least *some* cases of inhumane conduct do not show dehumanisation: "Their actions often betray the fact that their victims must seem human, all too human, to the perpetrators" (391, and again at 399, 400, 401, 403 and 404). Here, it is *because* the hated subject is sufficiently person-like (or human-like, in Manne's sense) that they are hated and mistreated. Manne explains this further as follows:

> "Under even moderately nonideal conditions, involving, for example, exhaustible material resources, limited sought-after social positions, or clashing moral and social ideals, the humanity of some is likely to represent a double-edged sword or outright threat to others. So, when it comes to recognising someone as a fellow human being, the characteristic human capacities that you share don't just make her relatable; they make her potentially dangerous and threatening in ways only a human being can be—at least relative to our own, distinctively human sensibilities" (Manne 2016: 399-400).

Whilst Manne takes humanness (or being sufficiently person-like, in our sense) to be an exclusively human trait, we can extend her arguments to the morally considerable robots discussed in Sect. 2. These robots are sufficiently person-like in the sense that they have relevant high-level properties (rationality, emotions, etc.), abilities, behaviours, skill sets, etc. In virtue of being sufficiently person-like, these robots can be viewed as dangerous or threatening to human-beings. Using Manne's arguments, it follows that this perception (that the robot is dangerous or a threat) could become increasingly virulent in nonideal cases (as in the cases in Sects. 4, 5, 6, 7, where robots were perceived to increase the risk of sexual violence and to contribute to job losses).

So, what does the above mean for robot rights? Largely, the proponents of robot rights views seem correct to suggest that robots ought to have rights when they are sufficiently person-like (however these person-like criteria are understood). If we accept that all morally relevant entities ought to have at least moral rights, then it follows that person-like robots (who are morally considerable, see Sect. 2) ought to also have these rights. However, as emphasised above, the current robot rights views seem to miss an important connection between a robot having rights and a robot also being an object of hate. In what follows, I will explain how the discussions of hate presented in this article (Sects. 3, 4, 5, 6, 7) could actually help the robot rights views, rather than undermine them.

First, robot rights views should explicitly state how the conditions for a robot being a rights-holder (being sufficiently person-like) are also the conditions that could allow a robot to also be an object of hate (see the above discussions of Manne and Sects. 3, 4, 5, 6, 7). We need to make

Footnote 29 (continued)

consciousness" (Levy 2009: 211). "My own argument in support of giving certain rights to robots is not that robots with consciousness should have rights *because* [sic] of their consciousness *per se* [sic], but that, because they have consciousness, such robots will be regarded by us in similar ways to those in which we regard other humans…for example by regarding those robots as having rights" (Levy 2009: 214).

"If automata were constructed with the capacity for human-level sentience, consciousness, and intelligence, everyone concerned with human rights should consider whether such entities warrant the same rights as those of biological humans" (Miller 2015: 370).

"If RAIs [robots and artificial intelligence] came close to replicating our general capacity for rational autonomy, there would be a case for according them a comparable moral status to human beings, with corresponding rights as well as responsibilities" (Tasioulas 2019: 69).

Conversely, Danaher (2019b) rejects this focus. He argues that robots could have moral status and rights if they are sufficiently similar to marginal humans (defined in terms of the cognitively impaired), and/or to nonhuman animals ['Defending Premise (2): What's the performative threshold?']. He argues that a robot will be sufficiently similar if it is 'roughly performatively equivalent' to a marginal human or a nonhuman animal. "This means that if a robot consistently behaves like another entity to whom we afford moral status, then it should be granted the same moral status" ('The Sophia Controversy and the Argument in Brief'). On this view, we take some criteria for moral status (e.g., emotions, feeling pain, etc.) and argue that if a robot can behave in the same way as entities that meet this criterion, then they should also be granted moral status.

[30] There are two main ways in which we could understand the claim "it is humans who determine the conditions for being a rights-holder". On the first understanding, human beings try to discover what it means for a being to be a rights-holder according to objective criteria (moral realism). On the second understanding, human beings more or less invent the story of what it means to be a rights-holder (moral anti-realism). Though either understanding could be defended, I will be adopting the first view (moral realism) for the remainder of this article. I thank an anonymous reviewer for encouraging me to clarify this point.

this statement explicit so as we can be aware of the potential biases and prejudices that we may (consciously or otherwise) have against robots. Note that I am not suggesting that proponents of robot rights (or anyone who discusses the rights of robots, whether in support or not), hate robots (intentionally or otherwise). Nor am I claiming that the connection between the conditions for being a rights-holder and being an object of hate mean that it is acceptable to deny rights to robots simply because we (could) hate them and want to destroy them. That would be akin to genocide or racism, and is obviously not a morally acceptable stance to take. Instead, my claim is simply that, by acknowledging this connection, proponents of robot rights views might be able to identify some of the biases and prejudices (unconscious or otherwise) that may prevent others from supporting robot rights. If we can identify the biases that cause people to have negative reactions towards robots (e.g., seeing them as eerie and threatening, see Sect. 6), and dispute these negative perceptions, then it may be easier to get the robot rights movement off the ground when the time comes (i.e., when we are satisfied that there are morally considerable robots who ought to have rights).

Second, by examining relationships characterised by hate, it may be possible for proponents of robot rights to identify specific human behaviours that robots ought to be protected against. For example, drawing on the arguments of Sects. 4, 5, 6, 7, we might suppose that morally relevant robots ought to be protected against destruction, or forced labour in dangerous environments, etc. In other words, by drawing on discussions of hate, we could learn what human threats robots are particularly vulnerable to, and what moral (and perhaps legal) protections they ought to have as a result. These considerations could help proponents of robot rights views to identify the scope of robots' rights (if and when they have these rights). As the Neely quote in Sect. 2 emphasised, it is important that we start engaging with these moral considerations now "before we are quite at the position of having such beings to contend with" (Neely 2014: 109).

9 Conclusion

This article has argued for two claims. First, humans could be in relationships characterised by hate with morally considerable robots. Second, it matters that humans could hate these robots. This is at least partly because such hateful relations could have long-term negative effects for the robot (e.g., by encouraging bad will towards the robots). The article ended by explaining how discussions of human–robot relationships characterised by hate are connected to discussions of robot rights. I argued that the conditions for a robot being an object of hate and for having rights are the same—being sufficiently person-like. I then suggested how my discussions of human–robot relationships characterised by hate (Sects. 4, 5, 6, 7) could be used to support, rather than undermine, the robot rights movement.

Acknowledgements This article has benefitted from numerous thought-provoking discussions with staff and students within the University of Birmingham's Philosophy department. Particular thanks to Jussi Suikkanen and Em Walsh for their detailed feedback on earlier drafts of this article. Thanks also to family and friends for their input on the initial ideas for this article, and to two anonymous reviewers for their helpful comments.

Compliance with ethical standards

Conflict of interests The authors declare no conflict of interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

AI AI-Tech UK (n.d) Emma the AI robot. https://ai-aitech.co.uk/emma-the-ai-robot. Accessed 6 Oct 2020.

Ben-Zeev A (2018) Is hate worst when it's fresh? The development of hate over time. Emot Rev 10(4):322–324

Bernstein D, Crowley K, Nourbakhsh I (2007) Working with a robot: exploring relationship potential in human-robot systems. Interact Stud 8(3):465–482

Borenstein J, Pearson Y (2010) Robot caregivers: harbingers of expanded freedom for all? Ethics Inf Technol 12:277–288

Bryson J (2010) Why robot nannies probably won't do much psychological damage. Interact Stud 11(2):196–200

Bryson J (2012) A role for consciousness in action selection. Int J Mach Consciousness 4(2):471–482

Cambridge University (2018) Evolving with the robots. Press release for Dr Hattice Gunes' 2018 Hay Festival Talk. https://www.cam.ac.uk/news/evolving-with-the-robots. Accessed 6 Oct 2020.

Campaign against sex robots (n.d) https://campaignagainstsexrobots.org/. Accessed 6 Oct 2020.

Campaign to stop killer robots (n.d.) https://www.stopkillerrobots.org/. Accessed 6 October 2020.

Castro-Gonzalez A, Malfaz M, Salichs MA (2013) An autonomous social robot in fear. IEEE Trans Auton Ment Dev 5(2):135–151

Cave S, Coughlan K, Dihal K (2019) 'Scary robots': examining public responses to AI. Leverhulme Centre for the future of intelligence.http://www.lcfi.ac.uk/resources/scary-robots-examining-public-responses-ai/. Accessed 6 Oct 2020.

Coeckelbergh M (2009) Virtual moral agency, virtual moral responsibility: on the moral significance of the appearance, perception, and performance of artificial agents. AI Soc 24(2):181–189

Coeckelbergh M (2010a) Artificial companions: empathy and vulnerability mirroring in human-robot relationships. Ethics Law Technol 4(3):1–17

Coeckelbergh M (2010b) Robot rights?" Towards a social-relational justification of moral consideration. Ethics Inf Technol 12(3):209–221

Coeckelbergh M (2014) The moral standing of machines: towards a relational and non-Cartesian moral hermeneutics. Philos Technol 27(1):61–77

Coeckelbergh M (2018) Why care about robots? Empathy, moral standing, and the language of suffering. Kairos J Philos Sci 20(1):141–158

Condliffe J (2019) This week in tech: Some workers hate robots. Retraining may change that. The New York Times. https://www.nytimes.com/2019/07/19/technology/amazon-automation-labor.html. Accessed 6 Oct 2020.

Danaher J (2017) Robotic rape and robotic child sexual abuse: should they be criminalised? Crim Law Philos 11(1):71–95

Danaher J, McArthur N (eds) (2017) Robot sex: social implications and ethical. MIT Press, Cambridge

Danaher J (2019a) The philosophical case for robot friendship. J Posthuman Stud 3(1):5–24

Danaher J (2019b) Welcoming robots into the moral circle: a defence of ethical behaviourism. Sci Eng Ethics 26:2023–2049

Danaher J (2019c) Building better sex robots: lessons from feminist pornography. In: Zhou Y, Fischer M (eds) AI love you—developments on human-robot intimate relations. Springer, New York, pp 133–147

DeCosta-Klipa N (2019) The CEO of Boston Dynamics says it 'really bothers' him when people call their robots terrifying. Here's why. Boston.com Online News, October 28; 2019. https://www.boston.com/news/technology/2019/10/28/boston-dynamics-robots-terrifying. Accessed 6 Oct 2020.

DeGrazia D (2002) Animal rights: a very short introduction. Oxford University Press, Oxford

Dennett DC (1994) The practical requirements for making a conscious robot. Philos Trans R Soc Lond Ser A Phys Eng Sci 349(1689):133–146

Fischer AH (2018) Author reply: why hate is unique and requires others for its maintenance. Emot Rev 10(4):324–326

Fischer A, Halperin E, Canetti D, Jasini A (2018) Why we hate. Emot Rev 10(4):309–320

Frank L, Nyholm S (2017) Robot sex and consent: is consent to sex between a robot and a human conceivable, possible, and desirable? Artif Intell Law 25(3):305–323

Gildea F, Richardson K (2017) Sex robots: why we should be concerned. The Campaign Against Sex Robots. https://campaignagainstsexrobots.org/2017/05/12/sex-robots-why-we-should-be-concerned-by-florence-gildea-and-kathleen-richardson/. Accessed 6 Oct 2020.

Gordon JS (2018) What do we owe to intelligent robots? AI & Society. 1–15.

Gordon JS (2020) Artificial moral and legal personhood. AI & Society. 1–15.

Groom V, Nass C (2007) Can robots be teammates? Benchmarks in human-robot teams. Interact Stud 8(3):483–500

Gunkel DJ (2018) The other question: can and should robots have rights? Ethics Inf Technol 20(2):87–99

Jaworska A, Tannenbaum J (2018) The Grounds of Moral Status. The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/grounds-moral-status/. Accessed 06 Feb 2021.

Kauppinen A (2015) Hate and punishment. J Interpersonal Violence 30(10):1719–1737

Kubinyi E, Pongrácz P, Miklósi A (2010) Can you kill a robot nanny?: Ethological approaches to the effect of robot caregivers on child development and human evolution. Interact Stud 11(2):214–219

Laukyte M (2017) Artificial agents among us: Should we recognize them as agents proper? Ethics Inf Technol 19(1):1–17

Lay S (2015) Uncanny valley: why we find human-like robots and dolls so creepy. The Guardian. https://www.theguardian.com/commentisfree/2015/nov/13/robots-human-uncanny-valley. Accessed 6 Oct 2020.

Levy D (2009) The ethical treatment of artificially conscious robots. Int J Soc Robot 1(3):209–216

Manne K (2016) Humanism: a critique. Soc Theor Pract Special Issue Dominat Speech 42(2):389–415

Marti P (2010) Robot companions: towards a new concept of friendship? Interact Stud 11(2):220–226

Mathur MB, Reichling DB (2016) Navigating a social world with robot partners: a quantitative cartography of the Uncanny Valley. Cognition 146:22–32

Matyszczyk C (2019) People hate competent robots, says study. ZDNet, March 14, 2019. https://www.zdnet.com/article/people-hate-competent-robots-says-study/. Accessed 6 Oct 2020.

Miller LF (2015) Granting automata human rights: challenge to a basis of full-rights privilege. Human Rights Rev 16(4):369–391

Miller O (2019) This sex robot can breathe using her 'AI chest cavity. Technowize Magazine, October 31, 2019. https://www.technowize.com/this-sex-robot-can-breathe-using-her-ai-chest-cavity/. Accessed 6 Oct 2020.

Mulvey B (2018) Can humans and robots be friends? Dialogue Univ 2:49–64

Neely EL (2014) Machines and the moral community. Philos Technol 27(1):97–111

Nyholm S, Frank LE (2017) From sex robots to love robots: Is mutual love with a robot possible? In: Danaher J, McArthur N (eds) Robot sex: social implications and ethical. MIT Press, Cambridge, pp 219–245

Prabhaker A (2017) The merging of humans and machines is happening now. Wired, January 27, 2017. https://www.wired.co.uk/article/darpa-arati-prabhakar-humans-machines. Accessed 6 Oct 2020.

Reggia JA, Katz GE, Davis GP (2019) Humanoid cognitive robots that learn by imitating: implications for consciousness studies. In: Chella A, Cangelosi A, Metta G, Bringsjord S (ed) Consciousness in humanoid robots. Frontiers in Robotics and AI, Frontiers Journal Series, pp 17–29.

Ryland H (2020) On the margins: personhood and moral status in marginal cases of human rights. PhD Thesis, University of Birmingham.

Sharkey A (2015) Robot teachers: the very idea! Behav Brain Sci 38:e65

Sharkey A (2016) Should we welcome robot teachers? Ethics Inf Technol 18(4):283–297

Sharkey N, Sharkey A (2010) The crying shame of robot nannies: an ethical appraisal. Interact Stud 11(2):161–190

Smith A, Anderson M (2017) Automation in everyday life. pew research centre. Last modified October 4, 2017. https://www.pewresearch.org/internet/2017/10/04/automation-in-everyday-life/. Accessed 6 Oct 2020.

Sorell T, Draper H (2014) Robot carers, ethics, and older people. Ethics Inf Technol 16(3):183–195

Sparrow R (2002) The march of the robot dogs. Ethics Inf Technol 4(4):305–318

Sullins JP (2011) When is a robot a moral agent. In: Anderson M, Anderson S-L (eds) Machine ethics. Cambridge University Press, Cambridge, pp 151–160

Szanto T (2018) In hate we trust: the collectivization and habitualization of hatred. Phenomenol Cogn Sci 19:453–480

Tasioulas J (2019) First steps towards an ethics of robots and artificial intelligence. J Pract Ethics 7(1):61–95

The Royal Society (2018) Portrayals and perceptions of AI and why they matter. http://lcfi.ac.uk/media/uploads/files/AI_Narratives_Report.pdf. Accessed 6 Oct 2020.

Tistelgren, M. (2018) Can I have a robot friend? MA Dissertation, Umea University.

Titcomb J (2016) Boston Dynamics' terrifying new robot endures bullying from human masters. The Telegraph. https://www.telegraph.co.uk/technology/2016/02/24/boston-dynamics-terrifying-new-robot-endures-bullying-from-human/. Accessed 6 Oct 2020.

Torrance S (2008) Ethics and consciousness in artificial agents. AI & Soc 22(4):495–521

van den Broek E (2010) Robot nannies: future or fiction? Interact Stud 11(2):274–282

van Doorn J (2018) Anger, feelings of revenge, and hate. Emot Rev 10:321–322

Whitby B (2010) Oversold, unregulated, and unethical: why we need to respond to robot nannies. Interact Stud 11(2):290–294

Young W (2019) Reverend robot: automation and clergy. Zygon 54:479–500

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.