International Journal of Social Robotics
https://doi.org/10.1007/s12369-019-00592-1

Designing Virtuous Sex Robots

Anco Peeters[1] · **Pim Haselager**[2]

Accepted: 11 September 2019

Abstract
We propose that virtue ethics can be used to address ethical issues central to discussions about sex robots. In particular, we argue virtue ethics is well equipped to focus on the implications of sex robots for human moral character. Our evaluation develops in four steps. First, we present virtue ethics as a suitable framework for the evaluation of human–robot relationships. Second, we show the advantages of our virtue ethical account of sex robots by comparing it to current instrumentalist approaches, showing how the former better captures the reciprocal interaction between robots and their users. Third, we examine how a virtue ethical analysis of intimate human–robot relationships could inspire the design of robots that support the cultivation of virtues. We suggest that a sex robot which is equipped with a consent-module could support the cultivation of compassion when used in supervised, therapeutic scenarios. Fourth, we discuss the ethical implications of our analysis for user autonomy and responsibility.

Keywords Sex robots · Virtue ethics · Human–robot interaction · Empathy · Narcissistic personality disorder

1 Introduction

Some may find it hard to come to grips with sex robots. Yet recent events, like the 2015 *Campaign Against Sex Robots* in the UK, the 2017 publication of John Danaher and Neil McArthur's volume on the ethical and societal implications of robot sex [18], and the fourth incarnation of the *International Conference on Love and Sex with Robots*, show that this topic has captured the public's eye and provokes serious academic debate. A recent report by the *Foundation for Responsible Robotics* [42] calls for a broad and informed societal discussion on intimate robotics, because manufacturers are taking initial steps towards building sex robots. We take up this call by applying virtue ethics to analyse intimate human–robot relationships.

Why should we look at such relationships through the lens of virtue ethics? Virtue ethics is one of the three main ethical theories on offer and distinguishes itself by putting human moral character centre stage—as opposed to the intentions or consequences of actions. Virtue ethics has been discussed in relation to artificial intelligence more generally [49,54]. However, virtue ethics has received relatively little attention in discussions regarding sex with robots, even though sex robots could have a significant impact on their user's moral character. Two main exceptions are Strikwerda [48], who assesses arguments against the use of child sex robots, and Sparrow [47], who suggests that rape representation by robots could encourage the cultivation of vices. Our aims are different, as we will not focus on either child sex robots or robots that play into rape fantasies. Instead, we propose how virtue ethics can be used to contribute to the potential positive aspects of intimate human–robot interactions through the cultivation of virtues, and provide suggestions for the design process of such robots.

We develop our thesis in four steps. First, we present virtue ethics in relation to other ethical theories and argue that, because of its focus on the situatedness of human moral character, virtue ethics is in a better position to assess aspects of intimate human–robot interaction (see also [51, p. 209]). Second, we show how our virtue ethical account fares better than current instrumentalist approaches to sex robots, such as those inspired by the seminal and pioneering work of Levy [32,33]. Such instrumentalist approaches focus too much on the usability aspects of the interaction and, unjustly, frame sex robots as neutral tools. Understanding the interaction with a sex robot as mere consumption insufficiently acknowledges

✉ Anco Peeters
mail@ancopeeters.com

[1] Faculty of Law, Humanities and the Arts (Building 19), University of Wollongong, Wollongong, NSW 2522, Australia

[2] Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

Published online: 23 September 2019

the risk of their influence on how humans think about and act on love and sex. Third, we propose a way to reduce the risks identified by considering how the cultivation of compassion as a virtue may help in practising consent-scenarios in therapeutic settings. This way, we aim to show how, under certain conditions, love and sex with robots might actually help to enhance human behaviour. Fourth, we examine the implications our virtue ethical analysis on intimate human–robot relations may have on our understanding of autonomy and responsibility.

2 Virtue Ethics and Social Robotics

Current ethical debates on human–robot interaction are generally not framed in terms of virtues, but in terms of action outcomes or rules to be followed. It strikes us as regrettable that up until now, virtue ethics has received relatively little attention in the literature on social robotics in general, and on intimate human–robot relations in particular (but see [3,27]). A virtue-ethical analysis can help evaluate how, on the one hand, human agents could make use of love and sex robots in ways that may be judged to be (un)problematic. On the other hand, virtue ethics may help to clarify how human behaviour and societal views are influenced by the use of such robots and thereby help us to learn more about what it is to be a virtuous person in an intimate relationship. To establish the potential of virtue ethics for the evaluation of intimate human–robot relationships, we will examine aspects of virtue ethics relevant to the current discussion and consider what it has to add compared to other ethical approaches.

Virtue ethics departs from the idea that the cultivation of human character is fundamental to questions of morality. In the Western philosophical tradition, Aristotle's theory of virtue ethics is the most influential and he defines virtue as an excellent trait of character.[1] Such traits, like honesty, courage and compassion, are stable dispositions to reliably act in the right way according to the situation one is in. Aristotle describes a virtue as, in general, the right mean between two extremes (vices). He states that courage, for example, can be described as the mean between recklessness and cowardice (*Nicomachean Ethics*, II.1104a7). Finding the right middle between extremes is a challenging task and approaching that middle often requires extensive practice. In addition to practice, acquiring a virtue is helped by instruction from a exemplary teacher. A virtuous person will have cultivated her character to be disposed to naturally act in the right way in the relevant situation. It should be noted that although virtues are not about singular acts, acting honestly, courageously or compassionately may help a person to become honest, courageous or compassionate. This potential interactive loop, of internalising behaviour by practice and feedback, motivates our interest in applying virtue ethics to intimate human–robot interaction.

Consequentialism and deontology are the two main rival theories to virtue ethics, and they dominate current discussions on the ethics of social robotics. Consequentialism is the ethical doctrine that takes the outcome of an action as fundamental to normative questions. Deontology or duty-based ethics takes the principles motivating an action as central to matters of morality. Operationalization of these frameworks can take different forms. For example, in the case of consequentialism, artificial agents could be programmed to evaluate the potential costs and benefits of an action [20,25,41,55]. Or, in the case of deontology, designers may strive to implement top-level moral rules in agents [19].[2] As consequentialism and deontology provide frameworks that can be translated relatively straightforward into implementation guidelines, they may be attractive from a roboticist's perspective. While we value the contributions of consequentialist and deontological approaches to the literature on robot ethics, we think that there are ethical issues which virtue ethics is in a better position to address. Such issues includes how, in the words of Vallor [51], advances in social robots are "shaping human habits, skills, and traits of character for the better, or for worse" (p. 211). Importantly, this insight supports the idea that robots are not neutral instruments, but that they may influence the way we think and act. We side, therefore, with other researchers who recognize that virtue ethics can be a fruitful framework for AI and robotics [3, p. 37].

There are at least three ways in which virtues (and vices) might play a role in social robotics. First, we may consider which virtues are or ought to be involved on the human side of robot design. For instance, is it desirable that a roboticist exhibits unbiasedness and inclusiveness when designing a robot? Second, robots may nudge users towards virtuous (or vicious) behaviour. An exercise robot, for example, can encourage proper exercise and discipline by giving positive feedback to its user. Third, robots may exhibit virtues (and vices) through their own behaviour. This can be illustrated by the Sociable Trash Box robot developed at Michio Okada's lab at Toyohashi University of Technology [56]: these robots exhibit helpfulness and politeness through their vocalisations

[1] Other influential virtue ethical traditions originated with, for example, Confucius or Buddhism. For reasons of space, we shall restrict ourselves to a (neo-)Aristotelian account of virtue, but we suspect that the investigation of other virtue traditions could yield an interesting intercultural approach to the ethics of social robotics. See also [51].

[2] Isaac Asimov's famous laws of robotics, often cited as illustration in the ethics of AI literature, are modelled after deontological formulations of how one ought to act. They brilliantly showcase the inherent tension between deontological robotic directives and the potentially disastrous consequences that strict adherence to these might have.

Fig. 1 The Sociable Trash Box exhibits helpfulness and politeness when it requests trash and then bows after receiving it. Reprinted by permission from Springer Nature: Springer *International Journal of Social Robotics* [56], © 2019

and bowing behaviour when they collaborate with humans to dispose of trash (see Fig. 1). So one could focus on the virtues of the designer, on the way robot behaviour affects the virtues of a human interacting with it, or on the virtues displayed by the robot, for instance, as an example to be followed or learned from. We will focus on the latter two points, but towards the end discuss their implications for design. We think it is likely that the degree of anthropomorphism [11,15,45,46] will play an important role for especially the second and third topics. This needs to be further investigated, but for the purposes of this paper we will discuss robots that tend towards the anthropomorphic rather than the more functional end—like conventional sex toys—of the anthropomorphism spectrum.

In relation to the third aspect, some have said that virtues might be difficult, or even intractable, to implement in a robot. This idea is motivated by the complexity of giving general, context-independent definitions of specific virtues and because an implementation of a virtue like honesty "requires an algorithm for determining whether any given action is honestly performed" [5, p. 258]. Although we acknowledge the specific implementation challenges that virtue ethics brings, we think these challenges can be addressed by looking at the underlying mistaken assumption that virtues need to be implemented top-down into the robot. Analogous to how humans learn to be virtuous not by being told what to do but by example, implementing virtues into the design of social robots can take a similar situational approach. For this reason, it has been argued that the "virtue-based approach to ethics, especially that of Aristotle, seems to resonate well with the […] connectionist approach to AI. Both seem to emphasize the immediate, the perceptual, the non-symbolic. Both emphasize development by training rather than by the teaching of abstract theory" [27, p. 249]. This resemblance, we suggest, can help inspire the implementation of virtues in modern-day robots. The use of machine learning with artificial neural networks may be a way of avoiding the need to write an algorithm that specifies what action needs to be taken when. Virtues that depend on, for example, recognizing emotions in a human and require an emotional response can be implemented by training a neural network on selected input—say, by analysing videos of previously screened empathic responses made by humans (as done by Refs. [29,31]). Through machine learning, robots could similarly learn to mimic certain behaviours that we might consider displays of virtue, such as a light touch on the shoulder to express sympathy. The challenging research question here would be how to operationalize this kind of training so that the robot learns from human teachers. Such implementations are not trivial, but they need not be intractable either.

Two potential points of critique need to be addressed before moving on. The first critique has been voiced by robot ethicist Robert Sparrow [47], who argues that sex robots could encourage vicious behaviour, while at the same time maintaining that he finds it hard to imagine sex robots could promote virtue. He proposes that if people own sex robots, they can live out whatever fantasies they have on the robots—even rape. He argues that repeated fantasizing and repeated exercise of potential representations of rape will influence one's character to become more vicious. Though we agree with Sparrow's premise that this development is problematic and deserves careful consideration, we disagree with the conclusion drawn. While rape representation might be facilitated by sex robots, this does not mean that the production of such robots need always be ethically inimical. Let us assume that rape-play between two consenting adults is not necessarily morally wrong.[3] What is potentially morally wrong in acting out this scenario, is that it might normalize the associated repeated behaviour outside of a consensual context—the cultivation of a vice. This could lead to unwanted degrading behaviour or generalization to other contexts involving human–human interaction. The same risk of inappropriate generalization applies to the scenario of the human–robot interaction. In the case of humans, this means that careful and continuous communication about what is allowed and what is not is crucial: the partners have to trust and respect each other in order to safely play out the fantasy and stay aware of the fact that it is a fantasy. Might a similar approach be possible to intimate human–robot interactions? We submit that there are ways to involve consent in the case of intimate human–robot interaction aimed to prevent the risk Sparrow is drawing attention to, without condemning the manufacture

[3] It is worth noting that on Sparrow's account one will have to bite the bullet and say that rape-play by consenting adults is morally wrong as well. Not everyone will be willing to accept this implication.

Table 1 Suppose we compare the multiple approaches in a hypothetical scenario where sexual consent is negotiated, verbal or otherwise, between two human partners. This table aims to showhowsuch a scenario can be analysed in the differentways discussed in the present paper. This rough distinction should not be taken to mean that, for example, consequentialism cannot talk about virtues. What distinguishes the different approaches is which concept they take to be central

	Consequentialism	Deontology	Virtue ethics	Instrumentalism
Fundamental concept	Action outcomes	Moral rule	Virtue	Instrumental use
Concept applied	Obtaining consent maximizes well-being for both parties	Obtaining consent is in accordance with the rule: "Do unto others as you would be done by"	Obtaining consent is compassionate and respectful	Obtaining consent is not necessary, unless required for obtaining satisfaction

and use of sex robots in principle.[4] It would require us to rethink sex education and the role sex robots can play in this, which we do in Sect. 4. Interestingly, if one accepts that sex robots may cultivate vices in humans, it seems possible that such robots potentially also cultivate virtues.[5]

A second issue that needs addressing is a more general critique against virtue ethics. It has been argued that virtue ethics as an ethical theory is "elitist and overly demanding and, consequently, it is claimed that the virtuous life plausibly could prove unattainable" [26, p. 223]. Why propose such a demanding ethical theory for framing human–robot interaction? First, because virtue ethics can do justice to an assumption we make, namely that intimate, sexual relations between humans and robots should be understood as bi-directional. In this context, bi-directional means that humans design robots, while the general availability of such robots in turn may influence human practice of and ideas on intimacy and love. In contrast, current ways of thinking about intimate human–robot relations often depart from an instrumental and unidirectional assumption. Such rival accounts understand these relations as the usage of tools by humans and see any influence that robots may have on humans as value-neutral. They are focused on the human perspective and therefore lose sight of important potential ethical implications of human–robot interaction, as we will argue in Sect. 3 and as illustrated in Table 1. Our assumption is in line with current developments in cognitive science and philosophy of technology, which suggest that the cognitive and moral dimensions of artefact interaction need to be understood from a distributed perspective that puts equal emphasis on agent and environment [17,22,52,53].

Another and possibly even more exciting reason to engage with virtue ethics, is that thinking about virtues in relation to robots might actually help to make virtuous behaviour more attainable. This might be done through the habit-reinforcing guidance of humans by robots designed to promote virtuous behaviour: either by robots nudging human behaviour directly or by robots exhibiting virtues themselves.

3 Contra Instrumentalist Accounts

Recent discussions on intimate human–robot relations are often informed by the work of Levy [32,33]. Levy argues that humans will have physically realistic, human-like sex with robots and feel deep emotions for and even fall in love with them. Although we laud the pioneering work Levy has done to open up sex and love with robots for serious academic discussion, we argue that his framework fails to properly account for the ethical and social implications involved.

Regarding sex, Levy suggests that, physically speaking, realistic human-like sex with robots will be possible in the near future. Though Levy paints a colourful history of the development of sex technologies, discussion of this is not of prime importance for our argument and we will not examine it further. For the present discussion, we will assume that the physical aspects of these robots can be worked out more or less along the lines which Levy describes. Interestingly, Levy goes so far as to say that "robot sex could become better for many people than sex with humans, as robots surpass human sexual technique and become capable of satisfying everyone's sexual needs" [32, p. 249].

Regarding emotions and love, Levy suggests that it is possible that humans can be attracted to and even fall in love with robots. Without going into too much unnecessary detail, his argument proceeds in four steps. First, Levy lists what causes attraction of humans to each other. Second, he considers how

[4] Obviously, the consent provided by a robot does not amount to legally binding consent, just like the rape of a robot would not constitute legal rape, for the simple reason that a robot is not a legal person and not a sentient being. Hence, we are discussing here the implications of a robot behaving in a certain way, not necessarily implying the existence of human-like cognitive, emotional states or identical legal status.

[5] Sparrow [47] finds it "much less plausible that sustaining kind and loving relationships with robots can be sufficient to make us virtuous" (p. 473). He acknowledges, however, that such a claim needs to be supported by an argument as to why virtues are to be held against a standard different from vices and that this is a topic for further discussion. We do not share his intuition, though we agree with his latter point and would furthermore like to add that more empirical data on how human–robot interaction influences human behaviour is needed—which is one of the motivations for the proposal in Sect. 4 of the present paper.

affective relationships between humans and pets develop, and, third, how such relationships develop between humans and their *virtual* pets. Fourth and finally, Levy applies his findings to human–robot relationships.

Through a careful examination of feelings of bonding and attraction in humans, Levy comes to the conclusion that humans will likely develop similar feelings of bonding and attraction for robots. A large role in this narrative is reserved for the human tendency to anthropomorphize artefacts (see [13,45]). He submits that "each and every one of the main factors that psychologists have found to be the major causes for humans falling in love with humans, can almost equally apply to humans falling in love with robots" [32, p. 128]. It seems that there are no major hindrances for humans to, at some point in the future, fall in love with their robot. We can, in principle, agree, with this conclusion and it furthermore looks like recent preliminary empirical evidence supports it [40].

Obstacles on the path towards the use of love and sex robots are deemed by Levy to be of a merely practical nature. The robots described are presented as taking care and recognizing the needs of their human partner—in terms of the feelings of bonding and attraction he listed earlier. On several occasions [32,34, pp. 219, 233] Levy compares sex with a robot to masturbation, and uses that comparison as a reason why robot-sex would prevent cheating on one's partner [p. 234]—like in the case of soldiers on a long-term mission. Moreover, Levy describes this perspective on sex as a kind of "consumption" [32, p. 242]. It is for this reason that we characterize accounts such as Levy's as 'instrumentalist.' Love and sex robots, on such accounts, are merely tools to be used or products to be consumed. However, we suggest that such an instrumentalist perspective could lead to practices that provide cause for concern. Also, we are not convinced that a purely instrumentalist use of sex robots would make many people "better balanced human beings" [p. 240].

A first concern is that framing robot-sex as consumption underestimates the potential impact the acceptation of love and sex robots will have on the way love and sex are perceived. Consider a world where your "robot will arrive from the factory with these parameters set as you specified, but it will always be possible to ask for more ardour, more passion, or less, according to your mood and energy level. At some point it will not even be necessary to ask, because your robot will, through its relationship with you, have learned to read your moods and desires and to act accordingly" [32, p. 129].

Why would people, when such partners are available, be content with any kind of relationship, emotional or sexual, that would not adhere to this standard of perfection? Access to these robots would make it tempting to view relationships as essentially one-directional need-catering and effortless, especially perhaps for adolescents who grow up with such access. This is not how love and sex at present needs to be or even generally is conceived, and it goes deeply against the conception of a relationship as existing between two or more equal persons. Seeing humanoid robots capable of emotional and sexual interaction as tools is like being in a relationship with a slave. There lies an important question at the core of this issue, specifically on whether there are ways of considering the relationship between human and robot that are not slave-like. However, this falls outside the scope of the current paper (though for a beginning of an answer to this question, see [15]). In any case, this comparison illustrates the extent to which Levy's framework is unidirectional, which is further exemplified by his comparison of robot-sex with masturbation. Masturbation, at least generally speaking, is a solitary enterprise, and does not reflect the reciprocal interaction that characterizes a typical sex encounter between two partners.[6] Precisely because robot-sex does not amount to either masturbation or sex between consenting adults, one needs to address its particular ethical implications.

The second worry is that the instrumentalist approach allows for downplaying the risk of addiction inherent in interacting with robots that can perceive and immediately cater to their partner's every need. Consider how Levy describes that "robots will be programmable never to fall out of love with their human, and they will be able to ensure that their human never falls out of love with them" and "your robot's emotion detection system will continuously monitor the level of your affection for it, and as that level drops, your robot will experiment with changes in behaviour until its appeal to you has reverted to normal" [32, p. 118]. This sounds like the perfect gambling machine, which constantly updates its rules according to its user's desires—though these robots are potentially far more addictive than any currently existing gambling machine. We think this issue is insufficiently addressed by instrumentalist approaches such as Levy's, because, if one thinks of robots as merely neutral tools, as he does, then any risk of addiction rests solely on the shoulders of the user and not on a robot or its designers. However, it is an open question whether this is how robot-sex will be experienced by human users (or their significant others). Rather, we suggest that robots are not merely neutral tools.

A convincing argument in this regard is provided by Verbeek [53], who argues that for instance an obstetric ultrasound is not merely a neutral tool, a 'looking glass' into the womb. Its use raises important ethical questions, like "What will we do when it looks like our unborn child has Down syndrome?" or social pressure such as "Why did you decide to let the child [with Down syndrome] be born, given that you knew and you could have avoided it?", or more general societal questions like "Is it desirable that ultrasonography

[6] This also illustrates that robot-sex is not or need not always be wrong. This would be as extravagant a claim as the suggestion that masturbation is always wrong.

leads to a rise of abortions because of less severe defects like a harelip?" [53, p. 27]. This shows that the use of obstetric ultrasound influences our moral domain. It is naive to think that using technologies would not shape our behaviour and societal practices. Instead, it is better to think about this shaping of behaviour while designing technology. Similarly, instead of seeing robots as neutral tools, we should acknowledge that, for instance, robots may evoke more emotions in us than other tools do, as Scheutz [39] suggests. More importantly perhaps, the design and use of intimate robots presuppose or establish certain practices concerning 'appropriate intimacy.' At the very least, these practices and their underlying assumptions should be elucidated.

Two conclusions can be drawn from the above account. First, humans and technologies should not be seen as separately existing entities, with technology providing neutral products for human consumption. Secondly, ethical analyses are not based on pre-given ideas or criteria, but need to re-evaluate how human-artefact interaction may be influenced or radically changed by new technologies. This means that stakeholders participating in the design of technologies have a responsibility both in considering how their products will shape human behaviour and reflecting on the ethical issues that may arise with the use of their product.

On this view, designers are "practical ethicists, using matter rather than ideas as a medium of morality" [53, p. 90]. In this framework there is room for the moral aspects of technologies in a pragmatic context, without it becoming a 'thou shalt not'-like ethics. A virtue-ethical approach is exactly what the topic of intimate relations with robots needs, because interacting with a robot as an artificial partner is, even more so than with a regular artefact, a relationship which intimately shapes our own dispositional behaviour and societal views as well. On first sight, Levy seems open to a more interactive view when he refers to Sherry Turkle, taking up her line of thought in saying that he "is certain that robots will transform human notions" including "notions of love and sexuality" [32, p. 15]. The way Levy discusses situatedness resonates with the notions that humans and technologies should not be seen as strictly separate entities and that certain concepts are not pre-given but arise out of interaction between humans and artefacts. Does that mean Levy has successfully anticipated critique along the lines we have set out? It does not.

Although Levy seems sensitive to the two notions mentioned, in practice it is merely a lip-service to interactive human–technology approaches. His instrumentalist treatment of human–robot relations deals with humans and robots in terms of isolated atoms with only a one-way connection between them, from user to robot, without any consideration of the larger reciprocal interactive effects on behaviour and social practices. He does not analyse robot-sex in terms of the structures and situatedness he earlier described. Any instrumentalist framework will focus on the human, subject side of things and portray robots as neutral artefacts to be used. What Levy describes is a trend of an increasing acceptation of robot sex, not how it would actually constitute or change (our conceptions of) sex or intimate relationships. Even if one agrees that masturbation is not cheating—an open question, likely to be influenced by many contextual factors—that does not necessarily mean that having sex with a robot will not be considered as cheating. An intelligent android functions on a distinctively different level of companionship than, say, a vibrator. More dramatically, if instrumentalist thinkers on the one hand argue that an intimate relationship with a robot is possible and imply that these kinds of relationships can be as intense and realistic as intimate relationships between humans, then they should agree that being intimate with such a robot, while in a relationship with someone else, could be construed as cheating. At the very least, one has to concede that robot-sex in such a scenario cannot simply be equated to masturbation. In other words, even assuming that one would find it hard to imagine someone being jealous about one's partner using a vibrator, one could still imagine jealousy plays a role when one's partner engages in sexual activities with a very human-looking and acting robot.[7]

The analysis we have given shows that instrumentalist approaches may leave crucial ethical considerations unaddressed. Notions of love and sex will be changed by the development of humanlike robots. But how will these notions change? If we can have sex robots which are "always willing, always ready to please and to satisfy, and totally committed" [32, p. 229], what will that do to the way we view relationships? An understanding of robot-sex not as instrumental, neutral use of tools, but as involving a reciprocal interaction between human agents, robots and their designers is required to develop adequate answers to questions such as these. This is where virtue ethics can provide a guide for evaluation of such interactions.

4 Consent Practice Through Sex Robots

In order to investigate how sex robots could make a positive contribution to human moral character, we draw on virtue ethics for ideas on how to cultivate virtues and connect those to insights from current empirical data provided by literature on robotics and psychology. Our aim is to avoid the problem of cultivating vices through repeated unnegotiated practice—

[7] The Swedish science-fiction television drama *Äkta människor* (*Real humans*, 2012) depicts an example of this when the relationship between Therese (Camilla Larsson) and her husband turns sour because he grows jealous of her 'hubot'—a humanoid robot capable of exactly the functions Levy discusses. This depiction is fictional of course, but the force of the story at least casts doubt on any outright dismissal of the possibility that humans will become jealous of robots.

such as illustrated by Sparrow. Indeed, well designed robots may create the possibility to actually improve attitudes and behavioural habits regarding sex. First, consider the human–sex robot rape play scenario again. Previously, we argued that what is problematic about this scenario is not the act between consenting adults itself, but the potential normalization of behaviour it could lead to. For instance, the human participant may become accustomed to immediate satisfaction of desires through the use of a human-looking object and might extend the involved behavioural patterns to objectify other humans.

One way of preventing unwanted behavioural patterns is by providing sex robots with a module that can initiate a consent scenario. Like consenting humans, a robot and its human partner will have to communicate carefully about the kind of interaction that will take place and the human will be confronted by the subject-like appearance and the behaviour of the robot. And like in a relationship between humans, this communication could potentially result in the robot sometimes not consenting and terminating the interaction. Such interaction with a robot might prevent the practice of unidirectional behavioural habits and a resulting increased objectification of other humans.[8] This consideration suggests that the potential psychological and behavioural benefits of a consent-module will make it at least worthy of investigation. One should notice too, however, that a consent-module may negatively affect the potential economic gains of sex-robot producers, a consequence that is not our main concern here. Second, there are potential benefits with respect to sex practice and cultural perception in general in the consent-module, namely in cultivating the virtue of compassion. Though we focus on compassion for the sake of limiting the scope of this case study, other virtues, such as respect, likely ought to play a role in consent-practice as well. We take compassion here as the ability to care for and open up to another person without losing sight of one's own needs and feelings. Virtuous displays of compassion strike the right balance between care for others and for oneself. Compassion can motivate a desire to help others and we take it to be related to, though distinct from, empathy (see [28]).

A robot equipped with a consent-module could potentially be used to investigate ways of improving consent practice in general. Often, partners communicate their willingness to engage in sex through nonverbal cues [14]. Yet, because nonverbal cues can be ambiguous, miscommunication can and does occur [2]. In response, some governmental institutions have advocated the need for active, verbal consent. The practice of active consent has been met by at least two problems. First, even verbal consent does not necessarily mean that a partner is freely engaging in sex, because, for example, social pressure or substance abuse may be involved [35]. Second, explicit consent has met with cultural resistance, as men and women generally believe discussing consent decreases the chance that sex will occur [30]. Still, active consent is seen as a crucial way of combating sexual assault and rape, for example, at college campuses [1,8,12]. There is a need to change perceptions and practice, especially by men [4], concerning healthy consent and sexual practices. Virtuous sex robots – supervised—might help facilitate a much needed cultural change in this regard by further investigating ways of navigating consent.

The advantage of using sex robots over traditional top-down education is that the robots can provide a kind of embodied training that helps adolescents in negotiating sexual consent. Interaction with a compassion-cultivating sex robot could raise awareness of how these scenarios could play out and alter behaviour through training. A sex robot which not only can practise consent scenarios with a human partner, but which can actually cultivate a virtue like compassion could potentially be used in sex education and therapy. A robot cannot suffer and so any moral harm during education or training will be minimized. It seems to us that compassion is a suitable virtue to be practised using sex robots in sex education and therapy. If successful in clinical trials, such robots can be used to support a change in perception and behaviour of consensual sex on a larger scale, and not just with adolescents.

One might be sceptical as to whether robots can facilitate a dependable long-term change in compassion—both in negative or positive ways. It seems reasonable not to judge this prematurely, as assessing the long-term effects of sexual human–robot interactions requires empirical investigation by sexologists and psychologists. A number of interesting experiments on the influence of social robots on human behaviour in more general terms, have been done in the lab of Nicholas Christakis. In one (virtual) experiment [43], humans were placed into groups which had to perform a task. Unknown to the participants, these groups also contained robot agents. The robotic agents were programmed to make occasional mistakes which adversely influenced group performance. This behaviour led to the human participants who collaborated directly with a robot, to become more flexible in finding solutions that benefited group performance. Similarly, a related experiment [50] reported that humans who collaborated on a task with robots which made occasional mistakes and acknowledged their mistakes with an apology,

[8] On the other hand, one might argue, as Sparrow does, that a non-consenting robot could potentially facilitate (the representation of) rape scenarios even more if the human partner ignores the robot's consent. We do not have a solution for that problem here (although, for example, a simple 'complete close-and-shutdown' routine might be an option), but it is a main reason why we later in this paper suggest to test this kind of human–robot interaction in a therapeutic setting first, as testing under supervision may give us new insights on how to potentially deal with issues such as these. In any case, we are not convinced that this argument is sufficient to not further investigate the potential benefits of consenting robots.

became more social, laughing together more often, and more conversational.

The design of virtuous sex robots requires thinking about a setting in which to test and apply them. A case study will give the constraints necessary for the design to be specific and feasible. We further think that building a robot which can operate in long-term intimate relations in general first requires at least building a robot which can operate on a smaller timescale with a specific target audience. Furthermore, it would be necessary to have the support of supervisors—next to the AI researchers which should of course also be involved—that have professional training in psychology or psychiatry. We therefore propose to start with testing virtuous sex robots in a therapeutic setting.

As the specific target audience or participants, we suggest to consider persons who have been diagnosed with a narcissistic personality disorder (NPD [6]) as the common medical understanding of NPD aligns well with the previously given definition of compassion. We propose to consider NPD patients who are already within a therapeutic setting, as this means that testing can be done in a controlled environment, under supervision of professionals in psychiatry, psychology, and sexology. The robot's design, testing and development beforehand should involve these same professionals, especially regarding the potential effects of a robot's refusal of certain kinds of interaction. The anticipated link with compassion can be found in the latest edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5). In it, narcissism is described as a "pervasive pattern of grandiosity, need for admiration, and lack of empathy" [6]. Nine indicators are listed for narcissistic behaviour, of which the third, fifth, and sixth are of special interest for us here. Respectively, those indicators are about the narcissist feeling special, being exploitative in social relations, and lacking empathy. If compassion as a virtue is the golden mean between two extremes, then it seems that the narcissist, who feels better than others and is self-obsessed, is at one extreme of the spectrum.[9] We would describe this extreme (or vice) as having the tendency to being overly involved with oneself. Hence, training the virtue of empathy and compassion would be most relevant for this focus group. Designing and evaluating a robot aimed at influencing the behaviour of persons is the most prominent, and challenging, task to be set. Though there is a lack of information on successful NPD treatments [21], there is some preliminary evidence that empathic treatments of those with NPD have positive effects [10].

Obviously, operationalizing our proposal requires careful testing before the possibility of actual use in training is even considered, as the care for patients and the safety of those potentially harmed by their conduct is paramount. One potential worry might be, for example, that people with narcissistic tendencies become more proficient in their manipulations. Therefore, professionals involved would need to closely monitor the patients and signal such possible undesired effects. These cautionary words notwithstanding, the potential support of compassionate robots for NPD treatments is in line with the aforementioned preliminary evidence [10] and worth further investigation.

The next step in making the robot ready to teach compassion is by training it to give basic responses to certain kinds of behaviour. As proposed before, this could be done by training it on recordings of how compassionate people respond to different kinds of (inappropriate) behaviour. This means the robot has to recognize at least one extreme on the compassion spectrum in terms of behaviour of its partner, and has to perform behaviour appropriate to what it observes. Figuring out what good identifiers of those extremes are and what responses work best will need to draw heavily on the expertise of the psychiatrists involved.

Compassion is considered here as the virtue which lies between the extremes of only caring about oneself, the narcissist, or of only caring about another person. That means that a robot designed to treat these kinds of disorders should be able to direct behaviour towards the middle of the spectrum, where there can be a healthy focus on both caring for oneself and caring for others. We suggest that it may be worthwhile to investigate whether and how such behaviours could be influenced by a compassionate robot. If this turns out to have promising results, work can be done on improving the design and expanding the use of such robots for other settings and for other groups of people.

5 Implications of Virtuous Sex Robots

We have striven to demonstrate that virtue ethics provides a useful framework for analysing the implications of sex robots, as well as for making recommendations for the design and application of such robots. We consider robot-sex as involving and supporting a reciprocal interaction between human agents and robots instead of as a form of uni-directional instrumental tool use. Applying virtue ethics led us to suggest a consent-module for sex robots that could support the development or strengthening of compassion in supervised, therapeutic scenarios. As such, sex robots may contribute to the cultivation of virtues in humans. However,

[9] In the spirit of virtue ethics, one could consider Dependent Personality Disorder (DPD) to be the other extreme on the compassion spectrum [6]:

> They are willing to submit to what others want, even if the demands are unreasonable. Their need to maintain an important bond will often result in imbalanced or distorted relationships. They may make extraordinary self-sacrifices or tolerate verbal, physical, or sexual abuse.

It would be interesting to investigate how love and sex robots could be relevant for training and therapy for members of this group as well.

virtue ethics does come at a price. In addition to its potential of providing an interesting perspective of the issues surrounding sex robots, it may also raise new problems. As an illustration of the latter, we would like to briefly reflect on two implications of implementing a consent-module. Robots saying 'no' towards the human that uses or owns them can lead to at least two related principled problems and one big practical challenge.

First, robots that refuse to comply with the demands or wishes of human beings may obstruct a person's autonomy, for example, as expressed by someone's immediate or long-term desires (see for a field study in the context of service robots for elderly [9]). Second, there is the threat of a responsibility gap. Finally, there is the practical challenge of how to design such a consent-module. We will offer some minor suggestions to address the latter at the end of this section.

We will illustrate the problem of a user's autonomy by considering a simple example in a different context. Imagine a beer robot, a simple system that keeps a stock of beers cooled and that brings one on demand. Obviously, at some point this might result in intoxication of the person demanding the beer. To what extent should a ('virtuous') beer robot be enabled to refuse the demands for another beer? Even though the consequences of intoxication may be bad for the persons themselves, as long as no one else or no one else's property is hurt, one might conclude that it is an expression of a person's autonomy to keep the beers coming. It is only or at least primarily in the context of negative effects for other persons or legal agents, that one could morally or legally preclude someone from having their wishes gratified. So, on the one hand, the human should be in control, but at some point or in certain contexts it could be legitimate or morally acceptable to limit the amount of control a human may have.

Regarding the responsibility gap, the problem is that when a human instructs a well-functioning robot to do something, and the robot is programmed to refuse to follow the instructions, all kinds of consequences may follow from that refusal for which the human, in essence, cannot or need not be held responsible. This leads to the question: Who would be responsible or accountable for any damages, psychological or physical, that may ensue? Of course, problems regarding the consequences of saying 'no' are not specific to virtue ethics. Rather, they are a consequence of any view that implies that robots under certain conditions should refuse specific instructions. However, this is worth discussing here because our analysis of virtue ethics leads to proposal of a consent-module, and its consequences should be noted. In our brief discussion, we will try to focus as much as possible on the specific nature of the ensuing problems in the context of sex robots.

In order to address these issues of autonomy and responsibility, we suggest considering the principle of 'meaningful human control'. This principle has been discussed in the contexts of military robots and self-driving. The principle states that ultimately humans should remain in control and carry (ultimate) responsibility for robot decisions and actions [7]. However, it is far from clear what this principle amounts to in practice, that is, what the requirements are for the robot so that it is capable of enabling this principle. de Sio and van den Hoven [44] indicate that humans merely 'being in the loop' or controlling some parameters may be insufficient for meaningful control if other parameters turn out to be more relevant to the robot's use or if the human lacks enough information to appropriately influence the process. In addition, possessing an adequate psychological capacity for (assessing) appropriate action is required for meaningful control, as is, thirdly, an adequate (legal) framework for assessing responsibility for consequences. Santoni de Sio and van den Hoven then analyse meaningful control in terms of Fischer and Ravizza's [24] theory of guidance control. Guidance control is realized when the decisional mechanism leading up to a particular behaviour is "moderately reason-responsive", meaning that in the case of good reasons to act (or not), the agent can understand these reasons and decide to act (or not), at least in several different relevant contexts. Moreover, the decision-making mechanism should be "the agent's own", in the sense that there are no excusing factors such as being manipulated, drugged, or disordered.

This, admittedly brief, consideration of meaningful guidance control provides a criterion that might be useful for the consent-module. It provides ground to think that when a human does not possess sufficient guidance control, or, by robot compliance with human instructions, may lose such control, a robot could be justified in non-compliance. This leads to two questions that need to be answered before a virtuous sex robot can be enabled with a consent-module, allowing it to refuse commands:

1. Is the person giving the current command in a state of meaningful human control?
2. Will complying with the current command lead to a reduction of meaningful human control, such that (5) is no longer the case?

In relation to the first question, the beer robot could make use of relatively reliable physiological measurements (like breath or blood analyses), or behavioural observations (like slurred speech or coordination difficulties). It will be more difficult to figure out which input patterns might engage the consent-module to generate refusals. Here too, the expertise of psychologists and psychiatrists, in relation to NPD for instance, is required. The main suggestion here is that a DSM-5 classified disorder in itself constitutes a reason for at least considering the possibility that the ability to act reasonably and compassionately might be affected, or that sound judgement and behavioural control might be impaired. Prac-

tically speaking, it would be relevant to investigate the extent to which data acquisition methods related to emotion recognition and sexual harassment might apply. Among potential indicators one could think of, for example, the human's lack of allowing turn-taking in communication, tone of voice and body posture, neglect of robotic non-verbal signals of non-interest, and so on (see, e.g., [36,38]). As a second step, investigations regarding the applicability of machine learning techniques are relevant (e.g., [23]).

The second question points to a difference between the case of the beer robot and the virtuous sex robot. In case of the beer, a prediction about the intoxication can be made on the basis of physiological variables. Given certain physiological aspects, the time course of the intoxication can be inferred with reasonable, and legally satisfactory, certainty. An intoxication level close to life-threatening alcohol-poisoning, just to mention a relatively clear case, could result in justifiable robot non-compliance. However, in the case of the virtuous sex robot such a prediction about the consequences of (non-)compliance is not as straightforward. For this reason too, it bears emphasis that we are suggesting the investigation of the consent-module within clinical contexts. Assuming, for the moment, agreement regarding the appropriateness of a robot's non-compliance in certain situations, there is still a further question about how the non-compliance should be put into effect. We just mention a few possibilities here. One option is that a robot may refuse to comply, provide an explanation in terms of its assessment of the potential negative consequences, and provide information aimed at improved self-understanding and self-control. Ideally, this could result in a retraction of the instruction given. Another option may be that the robot refuses and informs a support group of, say, significant others or therapists. A more extreme option would be that the robot refuses and stops functioning altogether, by way of an emergency close-and-shutdown operation. Finally, it is worth noting that we may need to stretch our concepts of autonomy and responsibility beyond the individual and recast them in terms of open-ended and ecological processes (see [16]). Unfortunately, picking up this topic lies beyond the scope of the present paper.

Undoubtedly, many other issues and ways of addressing them surround the notion of a consent-module. We have explicated the present ones to emphasize that virtue ethics does not provide easy solutions. Rather, it opens up a research domain in itself, one that comes with its own set of promises and difficulties that will need to be addressed.

6 Conclusion

The field of robotics advances rapidly and robot ethics ought to keep up. In the foreseeable future, there will be robots advanced enough to evoke, even if only for a few minutes, the experience in humans that they are interacting with another human being. Unless a ban is implemented [37], which we do not want to rule out, it is likely that love and sex relationships with robots will be formed. How can we best understand and evaluate such relationships? We have taken some initial steps towards answering this question by arguing that virtue ethics is better suited than instrumentalist approaches to evaluate the subtleties of intimate human–robot relationships. Next steps should involve careful testing and with this in mind we have outlined how testing a consent-module for robots in a therapeutic setting may yield useful insights. Importantly, implications for user autonomy and responsibility should remain in focus of future research.

Some challenges are anticipated. First, the misuse of sex robots could have a lasting impression on an adolescent learning about intimate relationships, but there is also a positive side to developing realistic looking and acting love robots. Such robots could train people how to behave confidently and respectfully in intimate relationships. In a therapeutic setting, such robots could be used to improve empathy or increase self-love in persons with respectively narcissistic or dependent personality disorders.

Another challenge is society's response to sex robots. It is difficult if not impossible to predict how our conceptions of love and sex will change with the introduction of love robots. One risk here is that a potential societal taboo on love and sex with robots would lead to fringe behaviours and scenes, similar to the domain of drugs and prostitution. It is therefore important that the topic of sex-robots, challenging, exciting, or revolting as it may appear to different parties, remains open for investigation and discussion.

The implications of developing love and sex robots are potentially huge and we have striven to tentatively chart one path, a virtue theoretical approach, within this domain. Advances in other robotic fields, like care robots or military robots, might have analogous implications. In these areas too, we should avoid the mistake of assuming that robots will not change the way we view healthcare and warfare. On the contrary, we need to consider and assess which of these changes would be desirable or should be avoided. In any case, we would do well to avoid the suggestion that all these developments are necessarily bad. We suggest that there is the possibility, worthy to be investigated, that some changes might be for the good. When we realize that the way we design and use such robots is bound to affect us, we can think about ways of improving ourselves through the technology, by careful consideration and monitoring.

Acknowledgements We dedicate this paper to our late colleague, teacher, and friend Louis Vuurpijl, who, with infectious enthusiasm, guided many students in their first steps into the field of robotics. Many thanks to Nick Brancazio, Miguel Segundo-Ortin, and several anonymous reviewers for their feedback on a previous draft of this paper.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Abbey A (1991) Acquaintance rape and alcohol consumption on college campuses: how are they linked? J Am Coll Health 39(4):165–169. https://doi.org/10.1080/07448481.1991.9936229
2. Abbey A (1991) Misperception as an antecedent of acquaintance rape: a consequence of ambiguity in communication between men and women. In: Parrot A, Bechhofer L (eds) Acquaintance rape: the hidden crime. Academic press, New York, pp 96–111
3. Abney K (2012) Robotics, ethical theory, and metaethics: a guide for the perplexed. In: Lin P, Abney K, Bekey GA (eds) Robot ethics: the ethical and social implications of robotics. MIT Press, Cambridge, pp 35–52
4. Adams-Curtis LE, Forbes GB (2004) College women's experiences of sexual coercion. Trauma Violence Abus 5(2):91–122. https://doi.org/10.1177/1524838003262331
5. Allen C, Varner G, Zinser J (2000) Prolegomena to any future artificial moral agent. J Exp Theor Artif Intell 12(3):251–261. https://doi.org/10.1080/09528130050111428
6. American Psychiatric Association (2013) Personality disorders. In: Diagnostic and statistical manual of mental disorders, 5th edn. American Psychiatric Association: Philadelphia. https://doi.org/10.1176/appi.books.9780890425596.dsm18
7. Article 36: Killing by machine: key issues for understanding meaningful human control (2015). http://www.article36.org/autonomous-weapons/killing-by-machine-key-issues-for-understanding-meaningful-human-control/
8. Banyard VL, Ward S, Cohn ES, Plante EG, Moorhead C, Walsh W (2007) Unwanted sexual contact on campus: a comparison of women's and men's experiences. Violence Vict 22(1):52–70
9. Bedaf S, Draper H, Gelderblom GJ, Sorell T, de Witte L (2016) Can a service robot which supports independent living of older people disobey a command? the views of older people informal carers and professional caregivers on the acceptability of robots. Int J Soc Robot 8(3):409–420. https://doi.org/10.1007/s12369-016-0336-0
10. Bender DS (2012) Mirror, mirror on the wall: Reflecting on narcissism. J Clin Psychol 68(8):877–885. https://doi.org/10.1002/jclp.21892
11. Björling EA, Rose E, Davidson A, Ren R, Wong D (2019) Can we keep him forever? Teens' engagement and desire for emotional connection with a social robot. Int J Soc Robot. https://doi.org/10.1007/s12369-019-00539-6
12. Borges AM, Banyard VL, Moynihan MM (2008) Clarifying consent: primary prevention of sexual assault on a college campus. J Prev Interv Community 36(1–2):75–88. https://doi.org/10.1080/10852350802022324
13. Breazeal CL (ed) (2002) Designing sociable robots. MIT Press, Cambridge, MA
14. Byers ES, Heinlein L (1989) Predicting initiations and refusals of sexual activities in married and cohabiting couples. J Sex Res 26:210–231
15. Cappuccio ML, Peeters A, McDonald W (2019) Sympathy for Dolores: moral consideration for robots based on virtue and recognition. Philos Technol. https://doi.org/10.1007/s13347-019-0341-y
16. Clark A (2007) Soft selves and ecological control. In: Ross D, Spurrett D, Kincaid H, Stephens GL (eds) Distributed cognition and the will: individual volition and social context. MIT Press, New York, pp 101–122
17. Coeckelbergh M (2012) Growing moral relations: critique of moral status ascription. Palgrave, Basingstoke
18. Danaher J, McArthur N (eds) (2017) Robot sex. Social and ethical implications. MIT Press, Cambridge
19. Danielson P (ed) (1992) Artificial morality: virtuous robots for virtual games. Routledge, London
20. Deng B (2015) Machine ethics: the robot's dilemma. Nature 523(7558):24–26. https://doi.org/10.1038/523024a
21. Dhawan N, Kunik ME, Oldham J, Coverdale J (2010) Prevalence and treatment of narcissistic personality disorder in the community: a systematic review. Compr Psychiatry 51(4):333–339. https://doi.org/10.1016/j.comppsych.2009.09.003
22. Di Paolo EA, Buhrmann T, Barandiaran XE (2017) Sensorimotor life: an enactive proposal. Oxford University Press, Oxford. https://doi.org/10.1093/acprof:oso/9780198786849.001.0001
23. Fernandes K, Cardoso JS, Astrup BS (2018) A deep learning approach for the forensic evaluation of sexual assault. Pattern Anal Appl 21(3):629–640. https://doi.org/10.1007/s10044-018-0694-3
24. Fischer JM, Ravizza M (1998) Responsibility and control: a theory of moral responsibility. Cambridge University Press, Cambridge
25. Floridi L, Sanders JW (2004) On the morality of artificial agents. Mind Mach 14(3):349–379. https://doi.org/10.1023/B:MIND.0000035461.63578.9d
26. Fröding BEE (2011) Cognitive enhancement, virtue ethics and the good life. Neuroethics 4(3):223–234. https://doi.org/10.1007/s12152-010-9092-2
27. Gips J (1995) Towards the ethical robot. In: Ford KM (ed) Android epistemology. MIT Press, Cambridge, pp 243–252
28. Goetz JL, Keltner D, Simon-Thomas E (2010) Compassion: an evolutionary analysis and empirical review. Psychol Bull 136(3):351–374. https://doi.org/10.1037/a0018807
29. Güçlütürk Y, Güçlü U, Baró X, Escalante HJ, Guyon I, Escalera S, van Gerven MAJ, van Lier R (2017) Multimodal first impression analysis with deep residual networks. IEEE Trans Affect Comput. https://doi.org/10.1109/TAFFC.2017.2751469
30. Humphreys TP (2004) Understanding sexual consent: an empirical investigation of the normative script for young heterosexual adults. In: Cowling M, Reynolds P (eds) Making sense of sexual consent. Ashgate, Farnham
31. Janssen JH, Tacken P, de Vries JGJ, van den Broek EL, Westerink JH, Haselager P, IJsselsteijn WA (2013) Machines outperform laypersons in recognizing emotions elicited by autobiographical recollection. Hum Comput Interact 28(6):479–517. https://doi.org/10.1080/07370024.2012.755421
32. Levy D (2007) Intimate relationships with artificial partners. Maastricht University, Maastricht
33. Levy D (2007) Love and sex with robots: the evolution of human–robot relationships. Harper-Perennial, New York
34. Levy D (2012) The ethics of robot prostitutes. In: Lin P, Abney K, Bekey GA (eds) Robot ethics: the ethical and social implications of robotics. MIT Press, Cambridge, MA, pp 223–232
35. Lim GY, Roloff ME (1999) Attributing sexual consent. J Appl Commun Res 27(1):1–23. https://doi.org/10.1080/00909889909365521
36. Miranda JA, Canabal MF, Portela García M, Lopez-Ongil C (2011) Embedded emotion recognition: Autonomous multimodal affective internet of things. In: Palumbo F, Pilato C, Pulina L, Sau C (eds) Proceedings of the cyber-physical systems workshop 2018, vol 2208. Alghero, Italy, pp 22–29
37. Richardson K (2016) Sex robot matters: slavery, the prostituted, and the rights of machines. IEEE Technol Soc Mag 35(2):46–53. https://doi.org/10.1109/MTS.2016.2554421
38. Rituerto-González E, Mínguez-Sánchez A, Gallardo-Antolín A, Peláez-Moreno C (2019) Data augmentation for speaker identification under stress conditions to combat gender-based violence. Appl Sci 9(11):2298. https://doi.org/10.3390/app9112298

39. Scheutz M (2012) The inherent dangers of unidirectional emotional bonds between humans and social robots. In: Lin P, Abney K, Bekey GA (eds) Robot ethics: the ethical and social implications of robotics. MIT Press, Cambridge, pp 205–222
40. Scheutz M, Arnold T (2017) Intimacy, bonding, and sex robots: examining empirical results and exploring ethical ramifications. In: Danaher J, McArthur N (eds) Robot sex. Social and ethical implications. MIT Press, Cambridge, pp 247–260
41. Sharkey N (2008) The ethical frontiers of robotics. Science 322(5909):1800–1801. https://doi.org/10.1126/science.1164582
42. Sharkey N, van Wynsberghe A, Robbins S, Hancock E (eds) (2017) Our sexual future with robots: a foundation for responsible robotics consultation report. Foundation for Responsible Robotics
43. Shirado H, Christakis N (2017) Locally noisy autonomous agents improve global human coordination in network experiments. Nature 545(7654):370–374. https://doi.org/10.1038/nature22332
44. Santoni de Sio F, van den Hoven J (2018) Meaningful human control over autonomous systems: a philosophical account. Front Robot AI 5:1–14. https://doi.org/10.3389/frobt.2018.00015
45. Sparrow R (2002) The march of the robot dogs. Ethics Inf Technol 4(4):305–318. https://doi.org/10.1023/A:1021386708994
46. Sparrow R (2016) Kicking a robot dog. In: 2016 11th ACM/IEEE international conference on human–robot interaction (HRI), IEEE, p 229, https://doi.org/10.1109/HRI.2016.7451756
47. Sparrow R (2017) Robots, rape, and representation. Int J Social Robot 9(4):465–477. https://doi.org/10.1007/s12369-017-0413-z
48. Strikwerda L (2017) Legal and moral implications of child sex robots. In: Danaher J, McArthur N (eds) Robot sex. Social and ethical implications. MIT Press, Cambridge, pp 133–152
49. Tonkens R (2012) Out of character: on the creation of virtuous machines. Ethics Inf Technol 14(2):137–149. https://doi.org/10.1007/s10676-012-9290-1
50. Traeger M, Sebo S, Jung M, Scassellati B, Christakis N (2019) Vulnerable robots positively shape human conversational dynamics in a human–robot team. Presented at Center for Empirical Research on Stratification and Inequality Spring 2019 Workshop at Yale University on January 31 (Unpublished manuscript)
51. Vallor S (2016) Technology and the virtues: a philosophical guide to a future worth wanting. Oxford University Press, Oxford
52. Varela FJ, Thompson E, Rosch E (1991) The embodied mind: cognitive science and human experience. MIT Press, Cambridge
53. Verbeek PP (2011) Moralizing technology: understanding and designing the morality of things. University of Chicago Press, Chicago
54. Wallach W, Allen C (2009) Moral machines: teaching robots right from wrong. Oxford University Press, Oxford
55. Winfield AFT, Blum C, Liu W (2014) Towards an ethical robot: internal, models consequences and ethical action selection. In: Mistry M, Leonardis A, Witkowski M, Melhuish C (eds) Lecture notes in computer science and advances in autonomous robotics systems, vol 8717. Springer, New York, pp 85–96. https://doi.org/10.1007/978-3-319-10401-0_8
56. Yamaji Y, Miyake T, Yoshiike Y, De Silva PRS, Okada M (2011) STB: child-dependent sociable trash box. Int J Soc Robot 3(4):359–370. https://doi.org/10.1007/s12369-011-0114-y

Anco Peeters is a doctoral student in Philosophy of Mind & Cognition and tutor at the University of Wollongong, Australia. He obtained degrees in Philosophy and Artificial Intelligence at Radboud University, Nijmegen, the Netherlands. For his doctoral dissertation, he works on comparing functionalist and enactivist approaches to mind–technology interaction.

Pim Haselager obtained master degrees in philosophy and psychology, and received the PhD in 1995 at the Free University of Amsterdam, the Netherlands. He is an associate professor (Theoretical Cognitive Science) at the Donders Institute for Brain, Cognition and Behaviour, at the Radboud University Nijmegen. His research focuses on the implications of Cognitive neuroscience and Artificial Intelligence for human self-understanding. He investigates the ethical and societal implications of research in, and the ensuing technologies of, CNS and AI, such as Robotics, Brain-Computer Interfacing, and Deep Brain Stimulation. He is particularly interested in the integration of empirical work (i.e., experimentation, computational modeling, and robotics) with philosophical issues regarding knowledge, identity, agency, responsibility and intelligent behavior.