Virtue and vice in our relationships with robots: Is there an asymmetry and how might it be explained?

Professor Robert Sparrow, Department of Philosophy, Monash University.

This is the pre-press version of the paper, which appeared as

Sparrow, R. 2020. Virtue and vice in our relationships with robots: Is there an asymmetry and how might it be explained? *International Journal of Social Robotics.* Published Online: 22 February: DOI: 10.1007/s12369-020-00631-2

Please cite that version.

ABSTRACT

In previous work, drawing on virtue ethics, I have argued that we may demonstrate morally significant vices in our treatment of robots. Even if an agent's "cruel" treatment of a robot has no implications for their future behaviour towards people or animals, I believe that it may reveal something about their character, which in turn gives us reason to criticise their actions. Viciousness towards robots is real viciousness. However, I don't have the same intuition about *virtuous* behaviour. That is to say, I see no reason to think that "kind" treatment of a robot reflects well on an agent's character nor do I have any inclination to praise it. At first sight, at least, this is puzzling: if we should morally evaluate some of our relationships with robots why not all of them? In this paper, I argue that these conflicting intuitions may be reconciled by drawing on further claims about the nature of virtue and vice and the moral significance of self-deception. Neglecting the moral reality of the targets of our actions is little barrier to vice and may sometimes be characteristic of it. However, virtue requires an exercise of practical wisdom that may be vitiated by failure to attend to the distinction between representation and reality. Thus, while enjoying representations of unethical behaviour is unethical, acting out fantasies of good behaviour with robots is, at best morally neutral. Only in the rare circumstance where someone might be forgiven for mistaking a robot for a real animal or person may spontaneous responses to robots be virtuous.

Virtue and vice in our relationships with robots: Is there an asymmetry and how might it be explained?

In previous work, drawing on virtue ethics, I have argued that we may demonstrate morally significant vices in our treatment of robots (Sparrow 2017). Even if an agent's "cruel" treatment of a robot has no implications for their future behaviour towards people or animals, I believe that it may reveal something about their character, which in turn gives us reason to criticise their actions. Viciousness towards robots is real viciousness. However, I don't have the same intuition about *virtuous* behaviour. That is to say, I see no reason to think that "kind" treatment of a robot reflects well on an agent's character nor do I have any inclination to praise it. At first sight, at least, this is puzzling: if we should morally evaluate some of our relationships with robots why not all of them? In this paper, I argue that these conflicting intuitions may be reconciled by drawing on further claims about the nature of virtue and vice and the moral significance of self-deception.

The structure of my discussion is as follows. Section I tries to show that there might be something wrong with "cruelty" to robots and thus that at least some forms of our behaviour towards robots should be subject to moral evaluation. Section II briefly outlines the popular "cruel habits" account of what might be wrong with mistreating robots and explain its limitations. In Section III, I introduce the idea of "virtue ethics" and suggest that it offers important resources to help us understand the ethics of our relationship with robots. Section IV then draws attention to the asymmetry in my intuitions about virtue and vice in our relationships with robots and argues that an explanation is required. In Section V, I draw on some general features of judgements about virtue and vice to offer an explanation of – and justification for – the asymmetry. Neglecting the moral reality of the targets of our actions is little barrier to vice and may sometimes be characteristic of it. However, virtue requires an exercise of practical wisdom that may be vitiated by failure to attend to the distinction between representation and reality. Thus, I conclude, while enjoying representations of unethical behaviour is unethical, acting out fantasies of good behaviour with robots is, at best morally neutral. Only in the rare circumstance where someone might be forgiven for mistaking a robot for a real animal or person may spontaneous responses to robots be virtuous.

Kicking a robot dog

In February of 2015 Boston Dynamics released a YouTube video to publicise their research, which featured a robot they called Spot (Boston Dynamics 2015). In order to showcase the capacity of the robot to regain its balance if it became unbalanced, the video included two scenes in which a man kicks the robot heavily in its "flank", whereupon the robot lurches sideways dramatically only to recover its footing. According to its YouTube page, the video has been watched some 20 million times and has been extensively commented upon. On reading through the comments, it is striking how many times people attest to feeling sorry

for the robot or to thinking that it was "cruel", perhaps even wrong, for the demonstrator to treat it the way he did. Kicking a robot dog is cruel, just as kicking a real dog would be cruel. This intuition is puzzling because, of course, the robot dog is not sentient and feels no pain.

Admittedly, not everybody shares the intuition that the "mistreatment" of the Boston dynamics robot was wrong. However, a brief consideration of three other scenarios suggests that more people are prepared to morally evaluate our treatment of robots than first appears.

First, imagine that the robot dog looked much more like a real dog and had the capacity to flinch, cower, and emit cries of pain. You see a man beating and abusing this dog, becoming more and more enraged, and eventually "killing" it by striking it with a baseball bat. It is hard to avoid the intuition that he is doing something morally reprehensible, even if he owns the robot and had bought it for this purpose (Bartneck et al. 2007b; Darling 2012).

Second, imagine a man who owns both a black and a white robot butler. This individual always speaks politely to the white robot and is gentle and considerate in his treatment of it. However, he orders the black robot around brusquely, often swears at it, and sometimes beats it viciously. I suspect that many people will think that this behaviour is both racist and wrong.[1]

Third, imagine that your ex-partner builds a realistic looking robot that looks like you, and then sexually assaults it. It's hard to avoid the thoughts that there would be something wrong in your ex-partner doing so and that this would remain the case even if you never became aware of it (Sparrow 2017).[2]

Both separately and together these three different imaginary scenarios suggest that we are, at least sometimes, prepared to morally evaluate people's treatment of robots and especially their cruelty towards or "abuse of" robots (Bartneck et al. 2007a; Bartneck et al. 2007b; Darling 2012; Rosenthal-von der Pütten et al. 2013). These judgements remain even when it is obvious that the robot itself is not sentient and when mistreating the robot doesn't violate the property rights of others.

The "cruel habits" argument and its limits

The first argument to which people typically turn in order to account for these intuitions is that the way we behave towards to robots is likely to translate to the way we behave towards animals and towards people (Darling 2012; Gutiu 2012). This argument has a long history and was most famously put forward by German philosopher Immanuel Kant, as an explanation of what might be wrong with cruelty to animals (Passmore 1975). In Kant's

[1] For evidence of the willingness of people to attribute race to robots on the basis of the colour of their surfaces, see Bartneck et al. (2018). For discussion of the implications of such a tendency for the ethics of our relationships with robots, see Sparrow (2019a) and Sparrow (2019b).

[2] It is natural to think that part of what might makes this wrong would be whatever upset it might cause you. However, if your ex's activity isn't (independently) morally wrong it's unclear why it should be so upsetting. Moreover, the claim that your being upset constitutes a reason to criticise your ex-partner's actions depends on such treatment of a robot being (independently) morally wrong.

ethics, animals don't "count" — they have no moral status and thus cannot be wronged — because they are not rational and therefore are not members of the "kingdom of ends". Nevertheless, Kant holds that cruelty to animals is wrong because — and in so far as — those who are insensitive to the suffering of animals are more likely to behave badly towards people (Kant 1996).

The cruel habits argument relies crucially on an empirical claim about the extent to which our behaviour towards entities that only refer to human beings or human behaviours rubs off on our treatment of real human beings. In the original version, it is the fact that the bodily movements and sounds produced by animals when they are mistreated resemble those of human beings when they suffer, which motivates the concern that insensitivity towards the former might lead people to become insensitive to the (real) suffering of human beings (Elton 2000). In the discussion about the ethics of mistreatment of robots, the empirical claim is that the way we treat robots shapes our behaviour towards the things (animals, people, individuals…) the robots represent. Similarly, critics of on-screen violence, pornography, or video games have often suggested that these media forms cause the behaviours they represent (Dines 2010; Bushman 2016). At the highest level, the empirical claim concerns the relationship between representation and reality or between fantasy and reality. We should be self-conscious about the nature of our engagement with literature, film, video games, and robots because the way we treat representations of things influences our behaviour towards the things themselves.

Given its centrality to some of the most controversial ethical and political debates of the last several decades, one might have thought that the empirical premise would have been thoroughly researched and its truth or falsity convincingly established. Yet the claim that our enjoyment of fantasy shapes our behaviour remains massively and bizarrely controversial (Commonwealth of Australia 2009; Jenkins 2012; Wright, Tokunaga, and Kraus 2016, p. 183). Essentially, both academic and popular opinion are split into two camps each convinced of its own perspective and scornful of the alternative. One camp, which includes many social psychologists and some feminists, holds it to be obvious that what we fantasise about influences our behaviour and, especially, that exposure to media representations of violence and/or sexism increases violent and/or sexist behaviour in people (Dines 2010; Bender, Plante and Gentile 2018; Bushman 2016; Hald, Seaman, and Linz, D 2014; Wright, Tokunaga, and Kraus 2015). The other camp, which contains within it many games studies and cultural studies scholars, claims that there is little or no evidence that our enjoyment of representations or fantasies of unethical behaviour makes us more likely to behave unethically (Diamond 2009; Ferguson and Hartley 2009; Ferguson and Kilburn 2009; Sherry 2007). In particular, this camp emphasises the lack of evidence of the impact of media violence or sexism on rates of violent or sexual offences at the national level.

The extent of the controversy surrounding the empirical claim on which the cruel habits relies means that this argument has limited utility when it comes to mounting any sort of critique of our relationship with robots. One portion of the audience is already convinced, while the other is unlikely to be moved. Moreover, any debate about an argument along these lines is guaranteed to follow a very familiar path. For this reason, even those who are

personally convinced by the cruel habits argument, or something like it, would be well advised to seek out alternative accounts of the ethics of treatment of, and our relationships with, robots.

Although in what follows I will be engaged in this project, it is worth observing in passing that the HRI community often makes use of the claim that interactions with robots shape behaviour to motivate research into, or applications involving, robots. For instance, if robots are going to be useful for educating people (Belpaeme et al. 2018; Miller, Nourbakhsh, and Siegwart 2008) it will have to be the case that whatever people learn from the robots will shape their future behaviour. Indeed, sometimes robotics researchers argue that the embodied nature of robots means that our interactions with robots have *more* power to shape our behaviour that our interactions with other forms of media (Darling 2012). Even those who claim that indulging our fantasies with robots makes it *less* likely that we will act out our fantasies in reality are committed to a claim about the causal powers of robots.[3] Thus, although for political reasons I believe it is worth looking for alternative accounts of the wrongness to cruelty to robots, giving up on the claim about the causal powers of robots is likely to be to the detriment of social robotics research in the longer term.[4]

Virtue ethics

In previous work, I have suggested that virtue ethics offers valuable resources for explaining what might be wrong with "mistreating" robots without relying on the cruel habits argument (Sparrow 2017); a number of other authors have made similar claims (Cappuccio, Peeters, and McDonald 2019; Coeckelbergh 2007; McCormick 2001; Patridge 2010; Sicart 2010). Virtue ethics is a tradition of philosophical and ethical thought, often drawing on the writing of Aristotle, which argues that the best way to lead an ethical life is to cultivate desirable character traits ("the virtues") and to try to rid oneself of undesirable character traits ("the vices") (Aristotle 1986; Russell 2013). Where other ethical theories tend to focus on actions, virtue ethics focuses on the character of agents. Before we can answer the question "what should I do", we must ask what sort of person we wish to be. When it does come to the ethics of actions, virtue ethics argues that we should ask how they flow from, or evidence, the character of the actor. What kind of person would do that? What does it say about someone that they would do that? What would a person who possesses the appropriate virtues do in the same situation? (Annas 2011; Hursthouse 1999)

The advantage of virtue ethics when it comes to the evaluation of the ethics of our relationships with, and treatment of, robots, is that, because virtues and vices are partially constituted by characteristic thoughts, emotions, and fantasies (Oakley 1992), we don't

[3] See, for instance, the remarks attributed to Ron Arkin in Hill (2014).

[4] Moreover, no matter what we are inclined to say about the implications of enjoying "violent" videogames or pornography for behaviour, it seems highly unlikely that our enjoyment of particular sorts of representations doesn't shape our behaviour in the real world *at all*. For instance, advertising functions by associating the fantasy of pleasure with representations of products that consumers then become more inclined to purchase. Companies that rely on our buying their products spend a lot of money on advertising precisely because it works.

need to make claims about the implications of the way we treat robots for the way we might treat the things that the robots represent. “Cruelty” to a robot may reveal us to be cruel just because only a cruel person would take pleasure in “torturing” a robot. The dispositions and the emotions are themselves sufficient to establish that the action is vicious.

A puzzling asymmetry

Virtue ethical theories, then, can provide a compelling account of what might be wrong about cruelty to robots. They can explain our intuitions about the scenario, described above, where someone is involved in kicking or beating or otherwise abusing a robot dog.

However, something puzzling occurs if we consider the opposite case, where someone is being “kind” to a robot dog. Imagine someone who is always nice to their robot dog, who remembers to pat it every day, and take it for “walks”, who cuddles it, gives it “treats”, and speaks to it in a kind voice. What should we say about this case?

Unlike the case of virtual cruelty, I have no intuition that such behaviour is morally admirable or “virtuous”. If one believed in a “kind habits” argument — that kindness to robots will lead to kindness to the things (humans, animals, et cetera) that robots represent – then it would seem plausible that we should approve of kindness to robots.[5] However, as discussed above, my interest is in what we should say about the treatment of robots without making reference to the causal claim. My intuition is not that one could not *become* kind by practising kindness with a robot but rather that “kindness” to a robot is itself not genuine kindness. Nor is this intuition unsettled by the acknowledgement that, in general, the virtues, as with the vices, are as much a matter of our dispositions to feel and think, as they are to act. Thus, for instance, kind people will tend to feel happy when they learn of other people being kind and to feel the desire to help someone even when they cannot. Although I recognise this, I still cannot find it in myself to think that someone who is “kind” to their robot dog is doing anything good or admirable.

It turns out, then, that my intuitions about virtue and vice are asymmetrical. I am much more willing to criticise behaviour towards robots that I am to praise it. Indeed, it is not clear to me that I would *ever* be inclined to praise someone on the basis of the way they treat robots. While people can demonstrate real vices through their treatment of robots, they are not, I believe, able to demonstrate real virtues.

I regret that I have not yet had the opportunity to conduct a formal experiment to determine how widely shared is this pattern of intuitions. However, my experience in presenting work on this topic at conferences and seminars suggests that it *is* reasonably widely shared. Moreover, my own intuitions here are of the sort that I am inclined to try to defend: it’s not that I just *happen* not to approve of “kindness” to robots despite being prone to criticising cruelty to them but rather that I think that this is how things should be.

[5] Some supposedly educational uses of robots already seem to draw on this idea. For instance, the idea that after looking after a robot pet, children will be better able to look after a real pet seems to require that we can learn kindness — and not just to avoid cruelty — by practising with robots.

We *shouldn't* recognise virtue in people's engagements with robots (and other representations) while we should recognise vice in them.

Insofar as my main aim in what follows is to try to explain and defend this set of intuitions, and the asymmetry within, the paper is primarily addressed to those who share them. However, insofar as a defence of these intuitions makes them more attractive to those who do not currently share them, the paper also serves as an argument for this way of thinking about virtues and advice in relation to robots and other representations.

Accounting for the asymmetry

There are, I believe, two ways in which one might account for this asymmetry, which ultimately reinforce each other. Both lines of argument began with observations about the role played by virtue (and vice) in human life.

1. The precariousness of virtue

A first attempt to account for the asymmetry begins with the observation that just as there are more ways to be sick rather than healthy, or for something to be ugly rather than to be beautiful, there are more ways to be vicious than virtuous (Aristotle 1986 [1106b35]). A putative example of virtue is more likely to be vicious or otherwise fraudulent than an alleged example of vice is to be virtuous. Moreover, unacknowledged virtue is less damaging for an individual and for the community than unrecognised vice.[6] For both these reasons, our judgements about character are properly more critical — in the sense of being suspicious — of virtues than they are of vices.

Our intuitions about virtue and vice in general, then, are asymmetric: we are swifter to condemn vice then we are to praise virtue. One way this plays out, I believe, is in relation to acts that fail to achieve their goals and, especially, in relation to acts that were never likely to achieve their goals. We judge attempted murder, for instance, to be nearly as bad as murder (Feinberg 1995). But we don't think that attempted beneficence is almost as good as actual beneficence. This is not to say that we don't admire those who attempt to save a life, for instance, to some degree even if they fail. However, the gap between our valuation of attempts and success is larger when it comes to virtue than to vice. At least some virtues, then, are hostage to fortune in a way that vices are not. Moreover, when an action was *never* likely to achieve its goal this asymmetry is even more pronounced. A bumbling and incompetent attempted murder is attempted murder nonetheless and nearly as bad as murder. Our assessment of the agent's motives, and therefore character, is, for the most part, unaffected by their hopelessness. However, bumbling and incompetence in an attempt to help someone is corrosive of the intuition that the agent's motives were admirable. The inadequacy of the means speaks to the presence — or, rather, the absence – of the motive.

[6] Not recognising virtue in an individual may mean that we fail to praise or appropriately honour them, but they are unlikely to cease to be virtuous because of that, nor is the community likely to lose the benefit of the exercise of their virtue. However, a failure to recognise vice in someone is likely to allow them to continue to be vicious, to the detriment of their character and of those around them.

This is because, as I shall explore further below, “practical wisdom” – an understanding of, amongst other things, the way the world works — is more central to virtue than it is to vice. The larger point, though, is that in order to reduce the risk of misrecognising vice as virtue we pay more attention to the connection between means and ends when it comes to the attribution of virtue.

The asymmetry between our intuitions about virtue and vice in our relationships with robots may therefore be accounted for by reference to this larger asymmetry. Cruelty to robots isn’t actually going to hurt the robots. Nevertheless, that it is always doomed to fail in that regard detracts little from it being cruelty: it still expresses and evidences a cruel disposition. However, the fact that there is no sentient creature that could (directly) enjoy or benefit from “kindness” to the robot reduces the extent to which it is appropriate to describe this as kindness even though it is indeed some evidence of a kind disposition.

This argument goes a long way towards explaining why we might respond to kindness and cruelty to robots differently. But it will not justify my thought that there is *nothing* virtuous in kindness towards a robot. For that, we need to look to a different feature of the virtues.

2. Practical wisdom

A second explanation for the asymmetry, which also serves as a support for the first, emphasises the key role played by the virtue of “practical wisdom” in all the other virtues. The concept of practical wisdom plays a key role in Aristotelian virtue ethics. According to Aristotle all exercise of virtue requires the exercise of practical wisdom: those who don’t possess the virtue of practical wisdom cannot be said to have any virtues (Aristotle 1986, 224 [1144b10-33]).

Precisely what, Aristotle believed, practical wisdom consists in remains a topic of ongoing philosophical disputation. Nevertheless, a number of things are clear and should be sufficient to guide us in the current enquiry. First, practical wisdom requires an understanding of the nature of the good life for human beings. Second, it involves the capacity to understand how to act — which ends to pursue — in a particular situation informed by this knowledge of the good. This in turn means that, third, it requires knowledge about how the world works: it requires common sense and a modicum of practical skill (Kraut 2018). Practical wisdom requires knowledge of empirical matters and means because, insofar as they are oriented towards realising the good life, the virtues are oriented towards action (even when they are demonstrated through emotional responses), which means that they are oriented towards the world. If we don’t understand how the world works we cannot act virtuously. Lack of practical wisdom, however, is no barrier to the exercise of the vices, even if the vicious person might be less successful in realising their goals because of a failure to understand how the world works.

An emphasis on practical wisdom can explain how kindness towards robots isn’t virtuous at all. Robots are not appropriate objects of kindness or cruelty as they don’t feel anything. Machines cannot benefit from kindness or suffer as a result of cruelty. A person who possessed practical wisdom would know that and would also therefore realise that

"kindness" towards robots does not realise the goals — the improvement of the welfare of humans and (perhaps) other sentient creatures — towards which kindness is oriented. In the absence of the exercise of practical wisdom, virtue is impossible and what would ordinarily count as the exercise of a virtue — the demonstration of a kind disposition — fails to do so. However, that cruelty towards robots is similarly misdirected does not mean that it cannot be real cruelty because vicious behaviour need not be guided by the exercise of practical wisdom.

The role of practical wisdom in the virtues means that virtues need to be oriented towards appropriate objects in a way that vices need not be. We can be kind to people or animals but not to robots.

There is, however, a subtlety to this argument, which deserves exploration and somewhat complexifies a proper account of vice and (especially) virtue in relationship to robots (and other representations). What disqualifies "kindness" to robots from being kindness is its orientation towards the robot, which, as argued above, is not an appropriate object of kindness. But if the thoughts and feelings that would ordinarily count as kind are evoked because the robot represents something then, at least sometimes, we may wish to say that they are oriented towards the thing that the robot represents rather than the robot itself. For instance, if a person winces when they see someone kicking a robot dog, that response may flow from their concern for dogs rather than a concern for robots. In that case, it *would* be evidence of a kind disposition. Immediate emotional responses to representations can be virtuous. Yet behaviours, including the choice to seek out particular representations (for instance, of people suffering or receiving kindnesses) so that one can experience or take pleasure in particular emotional responses, seem more problematic because they involve mistaking, or sometimes just substituting, representations for reality and thus a lack of practical wisdom. For instance, it is difficult to see how patting a robot dog could be an expression of affection for real dogs. Similarly, while someone who enjoys pouring over accounts of rape and murder in true crime books may be vicious, someone who seeks out and enjoys stories of virtue risks mistaking representations of virtue for the real thing. Because practical wisdom is essential to virtue, if the responses to — or behaviour towards — a robot (or other representation) seem delusional then they cannot be virtuous.

This account suggests that there *is* one circumstance in which "kindness" to a robot *might* actually be kindness and thus virtuous, which is when a person genuinely mistakes the robot for something that is an appropriate object of kindness and where it was reasonable for them to do so. If there is no lack of practical wisdom in a person's response to a robot then appropriate responses towards it might well be virtuous. I say "might" here deliberately because it remains open to us to adopt an "externalist" account of practical wisdom whereupon we assess what practical wisdom requires of us with reference to the facts rather than the information available to an agent. But it would be equally plausible to conclude that, if a person could not possibly have known that what they thought was a dog was in fact a robot dog, or perhaps just if their mistake was reasonable in the circumstances, practical wisdom would place no barrier in front of their responding to the robot as if it were real, and thus to their responses to it being virtuous.

We are a long way, though, from knowing how to build robots that it would be reasonable to mistake for animals or people more than momentarily in ordinary circumstances. For the most part, then, while people can be vicious in relation to robots they cannot be virtuous. Virtue requires practical wisdom and practical wisdom requires that we direct our kindness to creatures who might actually benefit from it.

Conclusion

Virtue ethics has much to offer those who want to morally evaluate relationships with robots without relying upon claims about the way we engage with representations shapes our behaviour or — even more implausibly — arguments about the "rights" of robots. In particular, virtue ethics can explain why we might sometimes feel that it's wrong to be "cruel" to or "mistreat" robots by pointing out that such behaviour may reveal an agent to have a morally significant character defect.

While an account of vice in our relationship with robots (and other representations) should be welcomed, I have suggested here that the idea that we could demonstrate *virtues* in our relationships with robots is much more problematic. This asymmetry is puzzling. Fortunately, paying proper attention to the nature of virtue and vice, and especially to the role of practical wisdom in the exercise of the virtues, can, as I have shown here, explain and justify the asymmetry. Only in the rare case where it might be reasonable for someone to mistake a robot for the thing that it represents might spontaneous expressions of emotion and concomitant actions be virtuous.

This conclusion is, perhaps, a mixed blessing for the field of social robotics. By revealing how our relationships with robots can indeed be morally significant it highlights the significance of the choices made by roboticists, which can shape our relationships with their creations. However, it also suggests that the efforts of engineers can at most reduce the risk that their creations lead us astray. Our fantasies about immoral behaviour can make us vicious but our dreams of virtue cannot make us virtuous.

Of course, if we conclude that interactions with robots (and other representations) *can* shape our behaviour towards people and animals, the argument may well go differently. Even if we thought that "kind" behaviour towards robots wasn't really kind, for the reasons I have discussed here, a case might be made for the design of robots that encouraged such "kindness" in order that people might become more likely to be kind in their interactions with creatures that could actually benefit from it.[7] Equally well, though, the design of such robots will be more fraught insofar as cruelty to them would presumably also encourage cruelty towards people and animals.

Regardless, given the extent of the hostility towards claims about media effects, an investigation of the potential of a virtue ethical framework through which to evaluate our

[7] Designing robots to influence the behaviour of those who interact with them would raise a number of ethical issues beyond the scope of my discussion here, especially where this involves deliberately deceiving users about the nature or capacities of the robots. See, for instance, Boden *et al.* (2017) and Sparrow and Sparrow (2006).

relationships with robots, of the sort I have conducted here, is worthwhile in order to allow conversations about the ethics of our treatment of robots to proceed until the controversy about media effects is resolved (if it ever is). By highlighting the asymmetry between virtue and vice when it comes to our relationships with robots, and then showing how it might be explained, I hope I have also shown how these conversations might also contribute to our understanding of ethics and of the nature of "the good life" more generally.

Acknowledgements

I would like to thank Massimiliano (Max) Cappuccio and Omar Mubin for the invitation to present a keynote at the *Interdisciplinary Workshop on Robots & AI in Society* at Western Sydney University in 2018, which formed the basis of the current manuscript; I owe especial thanks to Max for his patience while I completed it. Thanks are also owed to Justin Oakley, Michael Flood, and Christoph Bartneck for discussions and/or correspondence during the drafting process.

Compliance with Ethical Standards:

Funding: This study was supported by the Australian Research Council's Centres of Excellence funding scheme (project CE140100012). The views expressed herein are those of the author and are not necessarily those of the Australian Research Council.

Conflict of Interest: The author declares that he has no conflict of interest.

References

Annas J (2011) *Intelligent Virtue*. Oxford University Press, New York.

Aristotle (1986) *Ethics*. Penguin Books, Harmondsworth, UK.

Bartneck C, Yogeeswaran K, Ser QM, Woodward G, Sparrow R, Wang S, Eyssel F (2018) Robots and Racism. In *Proceedings of 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI '18)*. ACM, New York, NY, USA. https://doi.org/10.1145/3171221.3171260.

Bartneck C, Van Der Hoek M, Mubin O, Al Mahmud A (2007a) "Daisy, daisy, give me your answer do!" Switching off a robot. In *2007 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 217-222). IEEE.

Bartneck C, Verbunt M, Mubin O, Al Mahmud A (2007b). To kill a mockingbird robot. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction* (pp. 81-87). ACM.

Belpaeme T, Kennedy J, Ramachandran A, Scassellati B, Tanaka F (2018) Social robots for education: A review. *Sci Robot* 3(21).

Bender PK, Plante C, Gentile DA (2018) The effects of violent media content on aggression. *Curr Opin Psychol* 19:104-108.

Boston Dynamics (2015.) Introducing Spot. *YouTube*, Feb 9, 2015. https://www.youtube.com/watch?v=M8YjvHYbZ9w.

Boden M, Bryson J, Caldwell D, Dautenhahn K, Edwards L, Kember S, Newman P, Parry V, Pegman G, Rodden T, Sorrell T (2017) Principles of robotics: regulating robots in the real world. *Connect Sci* 29(2): 124-9.

Bushman BJ (2016) Violent media and hostile appraisals: A meta‐analytic review. *Aggr Behav* 42: 605-613. doi:10.1002/ab.21655

Cappuccio ML, Peeters A, McDonald W (2019) Sympathy for Dolores: Moral Consideration for Robots Based on Virtue and Recognition. *Philosophy & Technology*. https://doi.org/10.1007/s13347-019-0341-y.

Coeckelbergh M (2007) Violent computer games, empathy, and cosmopolitanism. *Ethics Inf Technol* 9(3): 219-231.

Commonwealth of Australia (2009) *Australian Government Attorney-General's Department Discussion Paper: Should the Australian National Classification Scheme include an R18+ classification category for computer games?* Attorney General's Department, Canberra

Darling K (2012) Extending legal rights to social robots. Paper presented at the *We Robot 2012 conference*, Coral Gables, Florida, April 2012. http://dx.doi.org/10.2139/ssrn.2044797. Accessed 22 April 2016.

Diamond M (2009) Pornography, public acceptance and sex related crime: a review. *Int J Law Psychiatry* 32(5):304-314.

Dines G (2010) *Pornland: how porn has hijacked our sexuality*. Beacon Press, Boston.

Elton M (2000) Should vegetarians play video games? *Phil Papers* 29(1):21-42.

Feinberg J (1995) Equal punishment for failed attempts: Some bad but instructive arguments against it. *Ariz L Rev* 37: 117-133.

Ferguson CJ, Hartley RD (2009) The pleasure is momentary… the expense damnable? The influence of pornography on rape and sexual assault. *Aggress Violent Behav* 14(5):323-329

Ferguson CJ, Kilburn J (2009) The public health risks of media violence: a meta-analytic review. *J Pediatr* 154(5):759-763.

Gutiu S (2012) Sex robots and roboticization of consent. Paper presented at the *We Robot 2012* conference, Coral Gables, Florida, April 2012. http://robots.law.miami.edu/wp-content/2012/01/Gutiu-Roboticization_of_Consent.pdf. Accessed 22 April 2016.

Hald GM, Seaman C, Linz D. (2014). Sexuality and Pornography. In Tolman D, Diamond L, Bauermeister J, George W, Pfaus J, Ward M (eds.) *APA Handbook of Sexuality and Psychology: Vol. 2. Contextual Approaches*. American Psychological Association Washington, DC, pp. 3-35.

Hill K (2014) Are child sex-robots inevitable? *Forbes.com*. Available via http://www.forbes.com/sites/kashmirhill/2014/07/14/are-child-sex-robots-inevitable/#1053601f2ddf. Accessed 4 April 2016

Hursthouse R (1999) *On virtue ethics*. Oxford University Press, Oxford

Jenkins H (2012) The war between effects and meaning: rethinking the video game violence debate. In: Buckingham D, Willett R (eds) *Digital generations: children, young people, and new media*. Lawrence Erlbaum Associates, Mahwah, pp 19-32.

Kant I (1996). *Lectures on ethics* (ed) Heath P, Schneewind JB. Cambridge University Press. New York

Kraut R (2018) Aristotle's Ethics. *The Stanford Encyclopedia of Philosophy* (ed.) Zalta EN. https://plato.stanford.edu/archives/sum2018/entries/aristotle-ethics/.

McCormick M (2001) Is it wrong to play violent video games? *Ethics Inf Technol* 3(4):277-287.

Miller DP, Nourbakhsh IR, Siegwart R (2008) Robots for education. In: Siciliano B, Khatib O (eds) *Springer handbook of robotics*. Springer, Berlin; Heidelberg, pp 1283-1301.

Oakley J (1992) *Morality and the emotions*. Routledge, London.

Passmore J (1975) The Treatment of Animals. *J Hist Ideas* 36(2):195-218.

Patridge S (2010) The incorrigible social meaning of video game imagery. *Ethics Inf Technol* 13(4):303-312.

Rosenthal-von der Pütten AM, Krämer NC, Hoffmann L, Sobieraj S, Eimler SC (2013) An Experimental Study on Emotional Reactions Towards a Robot. *Int J of Social Robotics* 5: 17-34.

Russell DC (ed) (2013) *The Cambridge companion to virtue ethics*. Oxford University Press, Oxford.

Sherry J (2007) Violent video games and aggression: why can't we find links? In: Preiss R, Gayle B, Burrell N, Allen M, Bryant J (eds) *Mass media effects research: advances through meta-analysis*. Erlbaum, Mahwah, pp 231–248.

Sicart M (2009) *The ethics of computer games*. MIT Press, Cambridge, MA.

Sparrow R (2017) Robots, rape, and representation. *Int J of Social Robotics* 9(4): 465-477.

Sparrow R (2019a) Do robots have race? *IEEE Robot Autom. Mag*, Early access.

Sparrow R (2019b) Robotics has a race problem. *STHV*, Published online July 28, 2019. https://doi.org/10.1177/0162243919862862.

Sparrow R, Sparrow L (2006) In the Hands of Machines? The Future of Aged Care. *Minds Mach* 16(2): 141-161.

Wright PJ, Tokunaga RS, Kraus A (2016) A meta-analysis of pornography consumption and actual acts of sexual aggression in general population studies. *J Commun* 66(1):183-205.