Is kicking a robot dog morally wrong, or morally permissible?

Anthony Wallace (z5206356)

October 28, 2022

1 Introduction

As the pace of technological advancement accelerates beyond the scope of current moral frameworks, ethicists and developers must navigate a minefield of uncertain moral dilemmas. This is especially apparent in the field of robotics, where robotic dogs such as Sony's AIBO series [Sony Group Corporation, 2022] and Boston Dynamics' Spot model [Boston Dynamics, 2022] are beginning to reliably mimic their animal counterparts. As these robots become more realistic, a question then emerges: Is kicking a robot dog morally wrong, or morally permissible?

This essay argues that this action is morally wrong. This essay also suggests that rather than attempting to justify this rationale using deontological or consequentialist morality, both of which present difficulty when dealing with robots, one can justify that the act of kicking a robot dog is morally wrong under a virtue ethics framework. This action is a cultivation of vice that leads to a vicious character.

We begin by examining the moral quandary of kicking a robot dog in closer detail, and why it is important to justify for or against this behaviour. We continue by discussing how different branches of normative ethics approach this problem and their respective potential caveats. Finally, we present a case for the application of an agent-based form of virtue ethics as proposed by Mark Coeckelbergh, and discuss how virtue ethics can circumvent issues around the intrinsic properties of robots [Coeckelbergh, 2021b].

2 Addressing the Problem

The abuse of robots and robot dogs may seem like a novel idea lacking real moral ramifications. However, if we consider the nuances that such an argument presents, we will observe wider implications for the moral consideration of non-human actors and the justification behind modes of thinking in normative ethics.

Most people intuitively believe that the act of kicking a robot dog is morally wrong and can be considered an act of cruelty. We can empirically observe people displaying distress

when, for instance, the four-legged robots of Boston Dynamics are regularly beaten for interference testing. However, when most people are asked to explain why they respond this way, they struggle to articulate a sound justification [Sparrow, 2016]. The pace of development in robotics and artificial intelligence underscores the importance of justifying why abuse toward robots is immoral. After all, as a robot does not feel pain like a real dog would, there is no apparent moral reason to care for the robot's well-being or consider the act itself as wrong.

Part of the difficulty that this problem presents is the representational relationship that a robot dog would have with a real dog, especially as its appearance and mannerisms begin to approach verisimilitude. Perhaps the aversion one feels towards cruelty to robotic animals comes from associating robots with their representational form? Or put more simply, one imagines that the intended harm is directed towards the animal itself and its robotic simulacrum. Using a linguistic framework, one can construct a conceptual framework of an object as a person, which one can then engage in moral and social thinking [Gunkel, 2012]. This *other*-oriented relational approach provided by Gunkel can help to demonstrate why a person may empathise with a robot or care for its well-being. However, this descriptive approach does not give a moral protection or rationale for why such abuse would be immoral [Coeckelbergh, 2021c]. Conversely, we have to concede that representation plays an important role in considering this act immoral, otherwise we risk incorrectly extending the moral consideration of violence to all inanimate objects.

Consider the application that such a question and the resulting approach would have on other non-human actors and our relationships with them. Other than the robot dog itself, this pattern of thinking can concern any entity that we do not assign direct moral standing, or entities without explicit intrinsic properties that are assumed when dealing with human morality. Typically, we observe how the anthropomorphisation of entities such as the robot dog causes a perception shift that skews understanding towards a familiar preconception, in this case a dog [Coeckelbergh, 2021b]. Normative ethics can sometimes shift to categorise human and non-human when engaging with moral dilemmas, but by introducing uncertainty around a non-human entity's moral status, we find ourselves breaking down these categories. In justifying why kicking a robot dog is morally wrong, we are challenging an anthropocentric bias that humans are the only beings for which interaction has moral value [Coeckelbergh, 2021c].

The justification for or against the action of kicking a robot dog in normative ethics then, is imperative to providing a moral framework for or against the abuse of robots and robotic animals.

3 Applying Normative Ethics

Having defined the issue and its various nuances and implications, let us examine the perspectives offered by the normative moral branches of deontology, consequentialism, and virtue ethics.

When discussing the problem of robot dog abuse, there is a direct connection to the deontological concept of the Kantian dog in demonstrating how the abuse of non-human entities is morally wrong. From a Kantian moral perspective, the act of kicking a dog does not breach any ethical duty towards the animal. The Kantian refutes the notion of the animal as a rational being, thus disqualifying it from direct moral standing [Kant, 1997]. However, committing this act does indicate an alignment towards cruelty to humans, violating the direct duty we have to ourselves to act rationally and with kindness. In this ethical argument, the Kantian provides a deontological justification against harming animals, but also invokes the concept of indirect moral standing as the animal derives a moral standing only from its relational relationship with humans [Coeckelbergh, 2021c]. However, this interpretation can be problematic as it brings into question the intrinsic properties of the robot. In making a claim about the Kantian dog, we are considering the dog's intrinsic ability to feel pain and to suffer when assigning it indirect moral standing. If this is applied to a robotic dog, we cannot reliably assume the causing of suffering (even if said dog can simulate expression of pain) and so we cannot say that direct duty to ourselves has been breached.

Now let us consider the utilitarian or consequentialist perspective. The utilitarian would naturally evaluate the consequences of kicking a robot dog, and would likely approach the same conclusion as the Kantian; cruelty towards robots would have the consequence of cruelty towards humans, thus making the act of kicking the dog immoral. Here, we begin to sidestep the issue of intrinsic properties (though not avoiding them entirely). The utilitarian focusses on the consequences of the action irrelevant to the subject. However, this is also a problematic conclusion - while this statement holds in terms of animal welfare, in media studies there is still debate around the causality between the simulation of immoral acts and the realisation of these acts [Sparrow, 2021]. As such, we wish to avoid any consequentialist evaluation in relation to the robotic dog as a simulacra of a real dog.

Consider then, the virtue ethicist's approach to the problem. As they are primarily concerned with moral character, virtue ethicists would class the act of kicking a robot dog as a vicious act, as the underlying motivation does not originate from any virtuous source. Breaking down this action, the actor sought to cause "harm" to the virtual animal, an action that only a cruel person would take pleasure in (regardless of true harm caused) [Sparrow, 2021]. Executing this action thus damages their moral character and indicates a viciousness that may manifest through habit. Here, we manage to avoid considering the intrinsic properties of the robot as the subject does not determine the morality of the action; only the intentions behind the act are relevant. Virtue ethics also avoids the problem of the

consequentialist argument, as we do not imply that there will be violent consequences based on the action; only that it reflects badly on the actor.

The application of virtue ethics in this scenario has the potential to provide a comprehensive framework for the moral justification of avoiding robot abuse; however, it is worth exploring some nuances that such thinking can bring to the issue.

4 Virtue Ethics as Justification

As outlined at the beginning, this essay seeks to outline a method of agent-based virtue ethics that expands upon the moral character argument in more detail.

It is therefore necessary to examine how virtue and virtuous acts are defined, especially in the context of greater society. Coeckelbergh presupposes a temporal and performative definition of how virtues and vices are cultivated through societal perception and practice - wherein we have to consider how vicious acts are reinforced through inter-generational trauma, societal prejudices, and cultures of abuse [Bourdieu, 1990]. Looking at our theoretical vicious actor, we may see an abuser in a position of power over the robot. It is also important however to consider how a society's hierarchical structure incentivises domination over others who are perceived as subservient and susceptible to abuse. These social dynamics are internalised and embodied by the vicious actor over time [Coeckelbergh, 2021b]. This approach has us redefine virtue and vice as both habitual actions undertaken by an individual with a clear intent and embodying societal values cultivated over time. These values inform us of what the good life, or *eudaemonia*, of virtue ethics looks like; this practical wisdom informs us if the action we are doing is just [Russell, 2010].

This practical wisdom connects to one of the more contentious concepts concerning a virtue ethics approach. This contention lies in whether there exists an asymmetry between virtue and vice, where vicious acts directed towards a robot dog cultivate vice but virtuous acts do not cultivate virtue. Sparrow posits that there is an asymmetry in virtue ethics, as a practical wisdom is required to know if an action is virtuous - however, this practical wisdom would intuit that the virtuous act directed towards a robot dog would exist in isolation. He acknowledges the precariousness of virtue in that virtuous acts are only considered as such with both intent and result. Yet, a vicious act harms moral character if there is either vicious intent or a vicious result; there are more methods of vice than virtue. Sparrow suggests that as the virtuous act produces no true benefit, kindness does not have an effect on the robot dog and is better directed towards creatures that can benefit from it [Sparrow, 2021]. This view is not universally held, as Coeckelbergh disputes the existence of the asymmetry. Coeckelbergh addresses Sparrow's points of vice being more perceived than virtue by applying a normative, rather than descriptive, approach. Virtue should be seen equally to vice in an idealistic sense, and virtue should be praised more often. He posits that robotic dogs and other social robots can exist to train virtue without requiring an explicit benefit to the subject [Coeckelbergh, 2021a]. In presenting a method of virtue ethics for the moral protection of

robots, we accept the belief that performing virtuous acts to a robot dog improves moral character and shows a disposition towards kindness to humans.

Understanding the justification for calling the action vicious and examining its application to robots provides a more robust ethical justification to protect robots, based on the moral character of the humans that interact with them.

5 Conclusion

In examining the original ethical quandary: *Is kicking a robot dog morally wrong, or morally permissible?*, this essay explored the implications and nuances of asking such a question by considering representational politics as well as ramifications for non-human actors. Then, we explored the effectiveness of deontology, consequentialism, and virtue ethics in providing robust justifications for or against this action. We finished by presenting a framework of agent-based virtue ethics to justify against the abuse of robots as a vicious act that harms one's moral character. This method avoids problematic approaches involving the intrinsic properties of robots or any potential future consequences. This provides a comprehensive moral protection to not only robot dogs but any entity/object with indirect moral standing. Through this approach, we can interpret most human-computer interactions through the lens of a relational dynamic where we care for robots as a method for the cultivation of virtuous traits.

Hence, we come to the conclusion: the act of kicking a robot dog is morally wrong not because of any harm towards the robot, but because you harm your own character.

References

[Boston Dynamics, 2022] Boston Dynamics (2022). https://www.bostondynamics.com/products/spot. [Website].

[Bourdieu, 1990] Bourdieu, P. (1990). *The Logic of Practice*. Stanford University Press, Stanford.

[Coeckelbergh, 2021a] Coeckelbergh, M. (2021a). Does kindness towards robots lead to virtue? a reply to sparrow's asymmetry argument. *Ethics and Information Technology*, 23:649–656.

[Coeckelbergh, 2021b] Coeckelbergh, M. (2021b). How to use virtue ethics for thinking about the moral standing of social robots: A relational interpretation in terms of practices, habits, and performance. *International Journal of Social Robotics*, 13:31–40.

[Coeckelbergh, 2021c] Coeckelbergh, M. (2021c). Should we treat teddy bear 2.0 as a kantian dog? four arguments for the indirect moral standing of personal social robots, with implications for thinking about animals and humans. *Minds and Machines*, 31:337–360.

[Gunkel, 2012] Gunkel, D. J. (2012). *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*. The MIT Press.

[Kant, 1997] Kant, I. (1997). *Lectures on Ethics*. The Cambridge Edition of the Works of Immanuel Kant. Cambridge University Press.

[Russell, 2010] Russell, D. (2010). *Virtue ethics, happiness, and the good life*, pages 7–28. Cambridge University Press. Publisher Copyright: © Cambridge University Press 2013.

[Sony Group Corporation, 2022] Sony Group Corporation (2022). https://us.aibo.com. [Website].

[Sparrow, 2016] Sparrow, R. (2016). Kicking a robot dog. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 229–229.

[Sparrow, 2021] Sparrow, S. (2021). Virtue and vice in our relationships with robots: Is there an asymmetry and how might it be explained? *International Journal of Social Robotics*, 13:23–29.