

University of San Diego

Digital USD

Philosophy: Faculty Scholarship

Department of Philosophy

2023

On Respect for Robots

Daniel Tigard

Follow this and additional works at: https://digital.sandiego.edu/philosophy_facpub



Part of the [Philosophy Commons](#)

On Respect for Robots

Abstract

We spend a lot of time with robotic and artificially intelligent (AI) technologies today. At the same time, it appears that we are growing more accustomed to interacting with AI and robots as if they were fellow human beings. Such trends have aptly brought about increasing ethical discussions concerning how we should treat technological devices. Do we owe robots some degree of respect, and how could respecting robots be justified? With this article, I put forward a new way of answering these questions. I invoke a revisionist account of Kant's ethics that amends the usual priority of dignity before respect (Sensen, 2009). Doing so allows us to see how we might have good reasons to maintain respectful relations with some AI and robotic systems.

Keywords

robot ethics, AI ethics, human-robot interaction, dignity, respect, Kant

Disciplines

Philosophy

Notes

Tigard, D. (2023). On Respect for Robots. *ROBONOMICS: The Journal of the Automated Economy*, 4, 37. Retrieved from <https://journal.robonomics.science/index.php/rj/article/view/37>

On Respect for Robots

Daniel W. Tigard  ¹ 

Abstract


We spend a lot of time with robotic and artificially intelligent (AI) technologies today. At the same time, it appears that we are growing more accustomed to interacting with AI and robots as if they were fellow human beings. Such trends have aptly brought about increasing ethical discussions concerning how we should treat technological devices. Do we owe robots some degree of respect, and how could respecting robots be justified? With this article, I put forward a new way of answering these questions. I invoke a revisionist account of Kant's ethics that amends the usual priority of dignity before respect (Sensen, 2009). Doing so allows us to see how we might have good reasons to maintain respectful relations with some AI and robotic systems.

Keywords: robot ethics, AI ethics, human-robot interaction, dignity, respect, Kant

Type: Article

Citation: Tigard, D. W. (2023). On Respect for Robots. *ROBONOMICS: The Journal of the Automated Economy*, 4, 37

¹ Department of Philosophy, University of San Diego, 5998 Alcala Way, San Diego, CA 92110, USA; email: dtigard@sandiego.edu

 Corresponding author



© 2023 The Author(s)

This work is licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0).

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

1. Introduction

We spend a lot of time with robotic and artificially intelligent (AI) technologies today. At the same time, and perhaps as a result of this development, it appears that we are growing more and more accustomed to interacting with AI and robots as if they were fellow human beings (cf. Darling 2016; Nyholm 2020). Such trends have aptly brought about increasing ethical discussions concerning how we should treat technological devices, and for what reasons. Do we, for example, owe robots some degree of respect, and should they be granted personhood or rights? How, if at all, can we think of AI as possessing agency, and should we hold such systems responsible for their behaviors? No doubt, a great variety of responses to these questions can be found in recent literature.

In an influential 2010 article, Mark Coeckelbergh makes a ‘relational turn’ in the ethics of technology. With it, he offers a novel account whereby moral consideration for intelligent robots is grounded in the ‘moral-social significance of appearance’, rather than the classical ethical theories, particularly deontology, utilitarianism, and virtue ethics (Coeckelbergh 2010). These more common theories, he shows, focus too much on tying moral status to individuals based upon ‘hard’ ontological properties, like consciousness or sentience. Importantly, in closing, Coeckelbergh qualifies his account with the remark that ‘it is still not entirely clear if a turn to (social) relations demands a ‘paradigm shift’ that completely abandons existing ethical and social theories or if it requires a more moderate revision of these theories’ (ibid., p. 220, italics in original). Along with David Gunkel’s persuasive accounts of ‘thinking otherwise’ (2012; 2018), Coeckelbergh’s relational approach provides a highly fruitful route for making newfound sense of moral status in AI and robotics. That being said, I take his qualification as a call to action, namely to explore how the relational view can be accommodated and perhaps bolstered with more moderate revisions of classical ethical theories.

With this article, I invoke a revisionist account of Kant’s ethics that amends the usual priority of key concepts like dignity and respect. Whether in philosophical literature or in foundational governmental documents—such as the UN’s Universal Declaration of Human Rights (1948)—it is commonplace to first identify dignity as an absolute inner value, and *then* invoke a corresponding duty to respect. However, this common understanding may well be mistaken and in need of significant revision. Found in Oliver Sensen’s careful analyses, the revisionary picture holds, in short, that Kant ‘neither grounds the requirement to respect others on an absolute inner value all human beings possess, nor does he advance such a value’ (Sensen 2009, p. 328). By applying this distinctive deontological framework—one that is arguably a more accurate interpretation of Kant’s view—to our interactions with sophisticated technological devices, I put forward a new way of thinking about respect for AI and robotics. And while mostly consistent with Coeckelbergh’s social-relational justification, the account here shows that the relational turn might require only revisions to our understanding of a classic ethical theory. These revisions allow us to see how we might have good reasons to maintain respectful relations with some AI and robotic systems. Accordingly, the account developed herein will defend such reasons, namely for maintaining *respectful relations*, but it will ultimately show that the Kantian justifications for respecting robots turn out to be rather self-centered—which, as I suggest, does not necessarily count against it.

In support of this agenda, the present article proceeds as follows. In section 2, I sketch the necessary background on Kant and I outline Sensen’s crucial clarification of key concepts, namely dignity and respect. In section 3, I apply the revisionist picture of Kant to our interactions with today’s AI and robotic systems, in order to critique three potential reasons to maintain respectful relations with AI and robots. In section 4, I explore further connections between the revisionist Kantian account and the relational approach to granting moral consideration. In section 5, I close by clarifying the implications of my account and offering several caveats.

2. Dignity as a Relational Property

First, it will be important to get clear on some of the basic ideas in Kant’s ethics and how they are commonly utilized in applied settings, such as medicine and technology. Although an in-depth coverage of Kantian literature

is not possible here, it is also not necessary, and for these reasons I will restrict my discussion to a brief account of two key concepts at stake here: dignity and respect.

Throughout many prominent writings on Kant's ethics, particularly concerning his *Groundwork for the Metaphysics of Morals*, it is widely accepted that for Kant there must exist an absolute inner value, namely the dignity possessed by human beings. For example, in a recent article, Schönecker and Schmidt (2018) provide strong textual support for their claim that dignity and value are indispensable and help to form what they call the 'ground-thesis'—specifically, the idea that 'rational nature exists as an end in itself and is the ground of the categorical imperative' (Schönecker and Schmidt 2018, p. 81).¹ Accordingly, dignity is thought to be a value property within human beings, prior to and independent from their relations between one another, and which serves as the crucial justification for the demand to treat humanity with respect. As stated in the famed 'Humanity Formulation' of the categorical imperative, you must 'Act in such a way that you treat humanity, whether in your own person or in any other person, always at the same time as an end, never merely as a means' (*Groundwork*: 429). Undoubtedly, this passage is one of the key depictions of what it means for us to treat each other and ourselves with respect. When we use others for our own ends, we fail to respect them as persons, as autonomous and free beings capable of setting their own ends; we violate their inner dignity. Following Sensen (2009; 2011), I will refer to the view of dignity as an absolute, non-relational property that grounds the duty to respect others as the *contemporary* paradigm, considering its common usage in contemporary Kantian scholarship and in modern political documents, such as the Universal Declaration of Human Rights.² I should also note that this contemporary deontological account is clearly of the sort that Coeckelbergh (2010) aptly considers too rigid to apply to robots, as it ties moral status to 'hard' ontological properties in individuals.

The central role of rationality and the notion of rational human beings as ends in themselves should be somewhat familiar even to readers with a philosophical background covering only the essentials of Kant. It is also here that we see how Kantian ethics frequently runs into difficulties, and perhaps undesirable implications, namely when we apply this line of thought without qualification to sensitive questions concerning the treatment of creatures other than fully rational human beings. In the longstanding debate over the permissibility of abortion, for example, it appears that fetuses might not (yet) deserve respect—or at least, we find a degree of justification for according them less moral consideration than adult humans—if the only criterion is that they are not (yet) capable of rational thought.³ Likewise, for non-human animals and for seemingly mindless objects in the natural world, it is a classic challenge to the strict Kantian reading of dignity to ask why, if at all, we should treat them with respect.⁴ Nevertheless, as Schönecker and Schmidt firmly state, 'Kant claims that *human beings* as autonomous beings are ends in themselves that possess dignity and value, and that this moral status to be an end in itself is the ground of the [categorical imperative]' (Schönecker and Schmidt 2018, p. 82, *italics added*). For the contemporary Kantian view, then, holding dignity to be a non-relational property effectively excludes AI and robotic systems from having a moral status of their own.⁵

As I have prefaced, however, there is an alternative interpretation of Kant's ethics, one which leaves ample space for the moral consideration of non-human animals, the environment, and perhaps today's technological systems with which we increasingly interact on a social basis. The foremost revision of our understanding to be made here is in how we view the concept of dignity. While many have seen it as an absolute, non-relational property in humans alone, Sensen's work provides refreshing clarity and indeed a comprehensive taxonomy of every use of the term 'dignity' throughout Kant's works.

According to Sensen, Kant's published writings contain 111 occurrences of 'dignity', and 'only eight relate dignity to worth or value' (Sensen 2009, p. 320). These eight instances are what gave rise to the contemporary paradigm; yet even these uses fail to show that Kant maintained dignity as a non-relational value in humanity. Instead, as Sensen's analysis shows, in the few cases where dignity refers to worth or value, Kant spoke of an

elevated position—conforming to the ancient Roman usage of the term—specifically, the ‘elevation of morality over other forms of behavior’ and the idea of ‘sublimity or the highest form of elevation’ (Sensen 2009, pp. 321–324).⁶ Importantly, of the 17 occurrences of ‘dignity’ in the *Groundwork*, Kant ‘neither uses the term in connection with the Formula of Humanity... nor does he present an argument for an absolute value of human beings’ in the Third Section where he justifies his moral views (Sensen 2009, p. 321). As such connections are precisely what *would* support the contemporary view, given that they are lacking, we see powerful reasons to think Kant’s work cannot be used to defend dignity as an absolute non-relational property, and that dignity is not what justifies our duty to treat others—even human beings—with respect.

Why, then, should we respect humans, and might this justification extend beyond humanity? What exactly is dignity if not an absolute inner value? Here I must briefly establish Sensen’s positive account, then I will proceed to apply the view to ethical issues in AI and human-robot interaction. Fortunately, some of the details are readily observable when we consider characterizations that oppose the contemporary paradigm. Unlike the common accounts that view dignity as a non-relational property, Sensen shows that a closer reading of Kant reveals dignity as fundamentally relational. The ancient Roman *dignitas*, as suggested, refers to ‘an elevated position or rank’ – it was ‘a concept of political life: it expressed the *elevated position* of the ruling class’ (Sensen 2011, p. 75). Granted, Sensen does not find that Kant employs this exact usage, what he calls the *archaic* conception, as it typically applied only to a few. Rather, the archaic view of dignity gave rise to what Sensen dubs the *traditional* paradigm.

As opposed to the contemporary view, on the traditional paradigm dignity refers to a relational property—being elevated—and it is not seen as absolute in Kant’s work. Sensen shows that dignity is context-sensitive and described as two stages: there is the initial stage, namely of being elevated; but then one can realize (or fail to realize) one’s initial dignity, specifically by using (or failing to use) one’s capacities for freedom and reason. For instance, in speaking of duties toward oneself, Kant says ‘they consist in his being conscious that man possesses a certain dignity, which ennobles him above all other creatures, and that it is his duty so to act as not to violate in his own person this dignity’.⁷ Here we see that dignity expresses a special relational position, namely of humans above others in light of capacities for freedom and reason. Crucially, it also becomes clear that dignity is focused upon *oneself* and our behavior *toward* others, rather than being a property to which we respond *in* others. This is why what justifies the demand to respect others is not the ostensible fact that *they* possess dignity. Instead, it is the capacities for freedom and reason that we humans possess that command us to show respect. Acting accordingly is how we come to realize *our* dignity. As such, one need not rely upon ‘hard’ ontological properties to see that we owe moral consideration to others.⁸

3. Respectful Relations with Robots

Admittedly, there is much to be unpacked from the foregoing section, and I cannot do justice to each detail here. In the interest of applying the revisionist view of Kant to our interactions and relationships with emerging technologies, at present I want to highlight two fundamental ideas to take forward from Sensen’s analysis. Firstly, dignity is a relational property, and it finds its expression in our unique capacity to behave morally toward ourselves and others. On its face, this idea excludes AI and robotic systems from the class of individuals who possess dignity, just as it excludes, for example, fetuses and non-human animals.⁹ However, secondly, dignity is not *supposed* to be something we find in *others* in order to justify treating them with respect, despite the common interpretation. If it were, we would naturally find it very difficult to understand why we should respect anything other than fully rational human beings. As I explained, this is one of the key challenges to the widely held contemporary Kantian account. Instead, as Sensen helps us to see, on the traditional view, ‘the primary focus is not on the dignity of others, but on the realization of one’s own dignity’ (Sensen 2011, p. 84). Accordingly, there is ample room for carving out justifications to respect a wider range of others than human beings alone.

These ideas represent what is arguably the most accurate interpretation of Kant's highly influential ethical theory. They encourage us to look beyond the rigid ontological properties of others in order to justify moral consideration, as authors like Coeckelbergh and Gunkel suggest. As such, the revisionist account of Kant allows us to retain key pieces of deontological ethics while supporting the social-relational approach. To further bolster this approach in light of the Kantian revisionism, in this section I outline and critique three potential reasons we might have to respect AI and robots, starting with a classic Kantian line of reasoning.

(1) Disrespect might globalize, indirectly via our character

Even among the most widely accepted readings of Kant, it is commonly held that we do in fact have a mechanism that effectively extends moral consideration to beings other than fully rational humans. That is, we can be said to have duties to a wide range of others—say, to non-human animals—since our treatment of them will impact our treatment of our fellow humans. In Kant's work and in a great deal of Kantian literature, these are known as 'indirect duties', and they appear particularly forceful in cases of humanoid robotics.

As numerous authors have recently observed, humanoid robots—especially those designed for social purposes—tend to elicit responses in us that resemble responses we would have in our interactions with fellow humans (cf. Duffy 2003; Breazeal 2003). We seem to naturally anthropomorphize various creatures, as well as inanimate objects, in the sense that we attribute to them humanlike characteristics, from physical features to intentions and emotional states. As a result, it becomes entirely fitting to examine how we should treat these 'others' and whether our treatment of them is for their own sake or for *ours*. This sort of inquiry forms a guiding question for Sven Nyholm, who asks in his recent work *Humans and Robots* (2020), 'How should human beings and robots interact with each other, given... people's tendency to anthropomorphize robots?' In his explanation of the Kantian line of thought, Nyholm explains the argument—notably put forward by Kate Darling (2016)—that 'if somebody spends a lot of time being cruel to robots, this might corrupt and harden the person's character' (Nyholm 2020, p. 184). While he agrees that this mechanism can provide reasons for abstaining from cruel treatment toward robots, Nyholm suggests an extended Kantian formula of humanity whereby we should 'treat the apparent humanity in any person (or robot!), never merely as a means, but always as an end in itself' (ibid., p. 187). This reasoning provides a more direct mechanism against mistreating robots than considering potential long-term character corruption. However, as Nyholm concludes, because robots lack mental properties, any moral duties to treat them with respect 'are not moral duties owed to these robots themselves' (ibid., p. 198). Instead, what grounds our duty to respect particularly humanoid robots is their apparent humanity and our duty to respect humans; and so, while perhaps more direct in terms of immediate impact, the approach nonetheless resembles that of indirect duties.¹⁰

As I noted at the outset, it seems clear that we spend a great deal of time with sophisticated technology today, and very soon, it is likely that many of us will interact regularly and perhaps socially with robots. Where one's interactions appear disrespectful or even cruel—for example, gratuitously destroying essential parts—the common worry is that this treatment will globalize and, in our interactions with fellow humans, we will likewise be inclined to act in disrespectful or cruel ways. But if we focus on our duties to those fully rational humans, the thought goes, it turns out that we do have a duty to behave morally toward a more diverse class of others, since our behavior toward that larger class plays a crucial role in the long-term formation of our character or, for Nyholm, more immediately in our treatment of humanity.

In his analysis of various positions on robot rights, David Gunkel (2018) weighs in on this reasoning, aptly noting that, although the argument may seem intended to justify respectful treatment for non-humans, 'it's really all about us'. As Gunkel says of the Kantian argument—and of similar thinking in virtue ethics—the 'rights of others, in other words, is not (at least not directly) about them. It is all about us and our interactions with other human beings' (Gunkel 2018, p. 151).¹¹ Ultimately, the attempt at justifying respectful treatment of non-humans—

whether speaking of dogs, trees, or humanoid robots—often comes down to assuring our respectful treatment of those who possess dignity, that is, humans alone.

Keeping in mind the revisionist Kantian account above, however, it appears that our duties toward creatures like non-human animals are not derived from the fact that *other* humans have dignity. Again, it is the fulfillment of our *own* dignity that places us under an obligation to not disrespect others. Accordingly, the Kantian revision can be seen as expanding the class of others whom we ought not treat poorly, but it seems to imply an even more self-centered rationale than the standard deployment of indirect duties. That is, the revision's focus on the realization of one's own dignity moves the concern for disrespectful character formation from 'it's all about us' to 'it's all about me.' In the following section, I will return to further discussion of indirect duties and Kantian revisionism, but first I want to sketch a second argument for why we might have reasons to behave respectfully toward technological systems.

(2) *Disrespect might globalize, directly via our actions*

The key idea here is similar to the worry that disrespectful behavior can negatively affect humans by going 'global' in our character. But here the concern is explicitly more direct and more immediate than the concern for the long-term development of one's character. In short, rather than potentially disrespecting other humans and ourselves via developing poor characters, we might face situations where acting in seemingly disrespectful ways toward a technological device can effectively—and much more directly—show disrespect to humans, namely through the action itself. For instance, consider an example offered by Gunkel in an article concerning robot rights in the *MIT Press Reader* (2020). Here Gunkel invites readers to consider a world where 'social robots, like Alexa, have a right to privacy for the purposes of protecting the family's personal data' (Gunkel 2020). Although the intent of this passage is mainly to show that there are different kinds of rights, and that granting one kind does not entail granting every kind of right, what is also noteworthy about this example is the justification for granting even this one right to a technological system. That is, just like the concern for developing a character that might behave disrespectfully toward humans, the idea of granting something like a right to a robot can again be grounded in assuring respectful treatment of humans, but much more immediately. As we can imagine, where a visitor in an Alexa-owner's house—whether an invited guest or a repairman—requests the Alexa search history for whatever reason, we might reasonably want Alexa to be equipped with the capacity to not obey such a command. Such a capacity could be implemented, surely technically speaking, but also legally speaking, namely by granting Alexa something like a right to privacy or non-disclosure.

Could this mechanism for granting robot rights, even if only of a limited and derivative sort, be applied to the concern for treating robots with respect? It seems that this move could be made in several ways. Just as we may understandably aim to protect humans from harm by granting something like a right to an AI or robotic system, as seen above, we might also aim to assure respectful treatment of humans by demanding that AI or robotic systems be treated with respect. Consider, for example, the robotic seal-looking pet known as 'Paro'. This robotic device was designed to aid elderly persons by providing a form of companionship and has been found to increase social interaction not only with the device itself, but also with fellow elderly home residents (Wada and Shibata 2006; Aminuddin et al. 2016). While conclusive research on the effects of using Paro is still forthcoming, as we can imagine, an elderly person may well grow attached to the device in similar ways one can reasonably grow attached to a biological pet, such as a dog or cat. In these cases, where someone behaves cruelly toward a robotic pet, the action itself—and not only the long-term effect on one's character—could be seen as morally wrong. An owner of Paro, for instance, could take offense at anyone treating her robotic pet in ways that appear disrespectful. In this way, although asserting a demand to treat the robot with respect may sound peculiar to some, it would be a welcomed development in human-technology relations, at least for the owners of beloved robotic systems.¹²

Returning to the application of the revised Kantian view, it seems that, while some may readily see and abide by the demand to act in ways that appear respectful toward robots, ultimately again the obligation centers around us as humans, or further, around the individual actor. That is, because dignity is relational on the revisionist account, it is not properly ‘located’ in others as a way of justifying respectful treatment, even when the treatment of our fellow humans is at stake and the impacts upon them are experienced directly and immediately. Dignity, as described above, is a two-fold context-sensitive property: we have our initial dignity, namely in virtue of our capacities for freedom and reason; we can then follow the dictates of morality and thereby fully realize our dignity. And so, again, acting in ways that show respect for others, including non-human animals and even robotic systems, may seem ever more plausible on the revisionist account. However, one must tread carefully on this line of reasoning, as it risks falling into an argumentative strategy where ‘it’s all about me’.

(3) *Disrespect might backfire*

For this third and final reason for why we might want to maintain respectful relations with robotics and other technological systems, I turn to an argument that from its outset is unapologetically self-centered. While at present the view is likely held among only a small population, there are at least a few who find it prudent to start worshipping AI. It is perhaps bizarre to some, but the reasoning is quite simple. Many expert opinions foretell the near-future arrival of a *superintelligence*—a system surpassing human intelligence in every way, perhaps synthesizing the powers of humanity with those of machines, and capable of rapid self-replication and improvement, often referred to as the ‘singularity’ (cf. Good 1965; Kurzweil 2005). Even in considering the mere idea of this phenomenon, it becomes reasonable to be concerned not only for how we should behave toward this sort of system, but also, and perhaps much more importantly, for how it might behave toward us.

This concern, along with the ethical attention to smaller scale AI applications, has been a driving force behind attempts to resolve the so-called ‘value alignment’ problem. How, if at all, can we be sure that a superintelligent AI system will behave in ways that promote our values, or even in ways that are consistent with our values? Can we somehow program human values into the AI’s source code, or could we teach a self-learning system—say, via positive and negative reinforcement—how to honor and abide by our values? These questions are famously addressed in Nick Bostrom’s *Superintelligence* (2014), which also puts forward some widely discussed worst-case scenarios. In short, a superintelligent AI might misunderstand our goals and values, or might take up its own goals, in such a way that leaves us powerless to stop it or to alter its trajectory. Whether or not it intends to bring about the demise of humanity, the actions of a superintelligent system could lead to our extinction; and so, some think, our best bet is to get on its good side as soon as possible. Since it will be essentially godlike in being all-powerful, and perhaps all-knowing, in an effort to preserve our existence we should worship AI as our new god.¹³

Again, the reasoning here may sound far-fetched or purely fictional, but it is worth taking seriously, at least for a moment, as a potential reason we may have to establish and maintain respectful relations with very sophisticated technology. Returning again to the revised Kantian account, we can ask: Would the reformed notion of dignity as relational help us in securing cordial interactions with a superintelligent system? Could respectful treatment suffice to appease an omnipotent and perhaps unintentionally destructive AI god? As I have assumed throughout, I want to continue supposing that even a superintelligent AI system would not itself possess *human* dignity. This seems true simply by definition, but also quite intuitive when we consider that, for Kant, human dignity is grounded in *our* capacities for freedom and reason. Even if the superintelligence possesses something like freedom in its decisions and actions, as well as something like rational thought processes, it seems safe to suppose that human dignity is indeed reserved for humans alone. And while it was precisely this view that led to difficulties in explaining why, if at all, we have reason to respect non-humans, here is where the revisionary account derives some of its strongest appeal. We do not need to find dignity in others in order to

justify respectful treatment. Dignity, on the arguably more accurate reading of Kant, is not an absolute inner value, as many have thought.

If dignity is understood as a context-sensitive relational property possessed by humans, where its full realization is attained by acting according to the moral law and showing respect for others, we might at least suppose that this mechanism plausibly extends to AI and robotic systems, just as it can extend to non-human animals. But again, would acting accordingly appease the all-powerful systems with which we interact? Would it help us to assure our continued existence? As I suggested, this potential reason for maintaining respectful relations with sophisticated technologies is unapologetically self-centered. More than the first two reasons, here we can see right through the initial façade of trying to justify respectful treatment of non-humans. Worshipping AI in an effort to avoid our own demise seems to be among the most blatantly anthropocentric practices we could engage in, at least in terms of human-technology relationships. I do not pretend to be in a position to assess whether or not appeasing godlike AI will work.¹⁴ One of the few things that can be confidently supposed, however, is that if the object of our worship truly possessed superintelligence, it would presumably become aware of the reasons for our worshipping practices toward it. Assuming it possessed the jealousy and spite of some of the other gods humans have worshipped, the practice of worshipping an AI god for our own sake would likely turn out to be self-defeating. Nonetheless, supposing that our respect or reverence for superintelligence were to support our continued existence, we would see reasons to continue engaging in such practices if only to assure our continued ability to act on the moral law and to exercise capacities for freedom and reason. In this way, then, respectful treatment of an AI god might be justified in terms of assuring respectful treatment—or *any* sort of treatment—of ourselves and our fellow human beings. In other words, just as our character development and immediate actions toward other humans provide indirect arguments for treating non-humans and robots with respect, we may likewise have indirect duties to respect superintelligent AI.

4. Kantian Revisionism and the Relational Turn

My examination of the reasons to establish and maintain respectful relations with AI and robotics has been rather exploratory. I have outlined three initial reasons, all of which turn out to be somewhat, if not extremely, self-centered. In this way, however, the account established here is set apart from some of the more prominent related works, namely those showing that respect for robots is grounded in their ‘apparent humanity’ (Nyholm 2020) or in our social relations with them (Coeckelbergh 2010; 2021). Indeed, rather than finding reasons to respect others via the demand to respect humanity or purely by way of relations, it appears that we can find good reasons to respect others—and effectively widen the class of others to be respected—by starting with oneself. Further, then, the application of the revised Kantian account seemed to exacerbate a sense of self-centeredness, since the revisionary notion of dignity risked replacing the already problematic idea that ‘it’s all about us’ (as humans) with the idea that ‘it’s all about me’ (namely the full realization of my own dignity). But is the ostensible self-centeredness a mark against Kantian revisionism and its application to robots? Does this view truly support social-relational approaches to moral consideration, and how is it distinct?¹⁵ In this final substantive section, I return to the discussion of indirect duties, and I address these questions. In particular, I suggest that the self-centered feature is not necessarily a mark against the application of the revised Kantian account to robots, that Kantian revisionism supports the relational turn, and that the revisionary Kantian view of dignity helps to reestablish the role of properties in ascribing indirect moral status.

First, as I have endeavored to show, applying the revised Kantian account to technological entities, or simply to non-human animals and features of the natural environment, is a way of expanding the circle of moral concern beyond human beings. On the revised Kantian view, as Sensen shows, dignity is relational and is not meant to be found in others in order to justify our moral treatment of them. Instead, when concerning ourselves with moral treatment and maintaining respectful relations, the focal point of dignity is that of the moral agent and not of the moral patient. In other words, dignity begins with the actor and not with the *other* toward whom (or toward which) we are acting. As outlined above, this view has the advantage of overcoming some of the

standard objections to Kantian ethics, namely since Kantian accounts often struggle to explain why we should respect anyone or anything other than human beings, since humans alone possess dignity. But on the revised view, again, dignity is not supposed to be found in others, as it is a relational property that begins with the moral agent.

Also outlined above is the idea that dignity comes in two stages, where one's initial dignity becomes realized or fulfilled when acting properly upon one's capacities for freedom and reason. And it is particularly here that we may reasonably worry that fulfilling one's dignity via respectful treatment of robots, or of non-human animals or even of humans, is far too self-centered and for that matter cannot be morally desirable. On one hand, it may be tempting to simply accept the self-centered focus of dignity and deny that because of this feature it is a morally undesirable approach. After all, just as we cannot be sure that others have minds, it seems unclear how to know when others have dignity; and so, starting the moral relation with a feature of oneself may be a way of adopting a precautionary approach to moral consideration.¹⁶ Again, it is important to bear in mind that, although dignity begins with moral agents, it is a relational property in the sense that one who possesses it is uniquely situated to treat others with respect. Relatedly, on the other hand, one might deny that this account is *overly* self-centered and maintain instead that it is precisely the focus upon oneself—one's use of reason, behavior toward others and so on—that carves out much needed justifications for treating a wide class of others with respect. In either case, the self-centered feature does not necessarily count against it.

Next, while it may be readily apparent, I want to briefly say how the revised Kantian account and its application to technological entities is at least consistent with a social-relational view and likely goes farther in the sense of providing positive support for such approaches. In his recent work on indirect moral status for robots, Coeckelbergh (2021) identifies and defends four conditions for giving some degree of moral standing to robots, specifically, conditions that pertain to humans as social, feeling, playing, and doubting beings. To summarize Coeckelbergh's list, it is proposed that robots should be ascribed moral status if: doing something bad to a robot renders someone a bad person in the eyes of others, a user develops feelings of attachment toward a robot, the robot participates in play or collaboration with humans, or the user has doubt about a robot's moral standing (*ibid.*, p. 5).

When any one of these conditions is fulfilled, and to the extent that any condition is fulfilled, Coeckelbergh argues, we should grant robots a degree of moral status—but again, these are indirect arguments for ascribing moral status. That is, in all four circumstances, we should think of robots as having moral standing, we have reasons to treat them in respectful ways, but not for their own sake. It is because we have duties to ourselves and to fellow human beings that we have a duty to treat robots respectfully, or at least to refrain from gratuitously disrespectful behavior toward them. Keeping Coeckelbergh's conditions in mind, it can be said that a person who gratuitously destroys a humanoid robot might develop a bad character and become more prone to disrespecting humans, as discussed above. One who behaves seemingly disrespectfully toward a robot that someone cares for appears to disrespect the person who cares for the robot. Disrespecting robots that play and collaborate with humans may well disrupt the (human) benefits of the play or collaboration. And in cases where we are uncertain, for example, in online interactions, disrespectful treatment of others may turn out to be disrespectful treatment toward humans (*ibid.*, pp. 6-9). For Coeckelbergh, all four of these kinds of interaction support indirect moral standing for robots, and they similarly support the relational turn, since the key mechanisms at stake are relations—whether human-robot or human-human relationships—and not a supposed property intrinsic in the other, namely the robot.

Importantly, Coeckelbergh explains, all of the proposed conditions for robots having indirect moral status are sufficient but not necessary. He states explicitly that 'there could be other reasons' and indeed there are additional reasons, including the idea that the starting point of dignity is within the moral agent. That is, on the revisionary Kantian view of dignity, we see reasons to respect others which are not grounded in a property of

the other. While on some readings of this interpretation, dignity may appear rather self-centered, as I have discussed, it is commonly accepted even in standard Kantian accounts that our duties to behave respectfully toward others can be traced back to our own dignity.¹⁷ In this way, we should feel the force of the demand to not gratuitously destroy a humanoid robot even where we are certain that it is merely a robot. Indeed, there will be cases where failing to respect robots will not reflect poorly on our character, where we would not thereby disrespect a human who cares for it, where we would not disrupt beneficial play or collaboration, and where there is no doubt that it is merely a robot with which we are interacting. Nevertheless, failing to maintain respectful relations with some robots may be a failure to exercise our capacities for freedom and reason.

In cases of human-robot interaction, it is likely widely agreed upon that it is the humans involved who maintain the ‘elevated’ moral standing. That is, of the two sorts of being, the human bears initial dignity. For this reason, it is the human who can be expected to behave in a dignified manner, regardless of the properties of the other. Unless robots can one day be said to have dignity of their own, failing to act with dignity toward robots will not be a failure to act respectfully for the robot’s sake. Nonetheless, gratuitous destruction of a robot—and likely many milder forms of disrespectful interaction—can be a failure to fully realize our own dignity. In this way, even in one-on-one human-robot interactions, where the interests of other humans are not at stake, we see reasons to avoid disrespectful behavior toward robots. In short, it is for our own sake that we respect our robot others. With this, we can add to our understanding of how robots can be granted indirect moral standing. Likewise, we see how the revised Kantian account supports the relational turn, namely by providing another rich framework and classic theoretical foundation whereby what is more important for moral treatment than the properties of the other is the relation that obtains.

Finally, however, precisely because the revised Kantian account and its application to robots provides yet another sufficient condition for indirect moral standing, and because of its close consistency with the established social-relational approach, in closing I should clarify its distinctiveness and importance. To be sure, my application of the revisionary Kantian view to AI and robotics shares with social-relational approaches the rejection of ‘hard’ ontological properties of others as the grounding for moral standing. Further, the application seen here can be seen as echoing and substantiating what seems to be a passing suggestion in Coeckelbergh (2021, p. 21), namely that it ‘could be argued that “direct” moral standing is relational’. After all, if the starting point of human dignity is oneself and dignity is relational, then it may well appear that the direct moral standing of other humans is ultimately grounded in our relations to them, or at least in the recognition that they possess their own initial dignity, which again is relational. In other words, what gives human beings—whether speaking of ourselves or others—direct moral status is a relational property.

For that matter, however, the account here parts ways with some instantiations of the relational turn, namely those that are committed to ontology as a secondary concern. Here I have in mind Emmanuel Levinas (1969), who, according to Coeckelbergh (2021, p. 19), argues against ‘establishing moral consideration on the basis of properties of individuals.’ Granted, there is a notable appeal to such an account in that it encourages us to focus more on our relationships and experiences with others than on what makes them who (or what) they are.¹⁸ This account finds additional appeal in Gunkel’s (2012, p. 151) analysis, where it is said that, according to Levinas, ‘Western philosophy has been the reduction of difference... by finding below and behind apparent variety some common denominator that is irreducibly the same.’ Yet, the revised Kantian account deployed here shows us an anomaly to this view on the usual method of ethics. In short, we need not reject properties as justifications for moral standing. In fact, as established above, the social-relational approach to attributing moral status can, and I believe should, readily embrace the use of *some* properties, since as we have seen, properties such as the dignity of moral agents succeed in affording moral status to a wide array of others.

As the revised Kantian account shows, social-relational approaches can invoke properties as reasons for respecting others, so long as those properties are understood as relational. Indeed, on the revised Kantian view,

dignity is understood as non-absolute, in the sense that it can be realized but, depending upon the relations that ensue, one might fail to fully realize one's dignity. It should be understood as context-sensitive, in the sense that it encourages one to attend to their unique positioning and relations to others with whom (or with which) they interact. And finally, as shown throughout, dignity is properly understood as relational. No doubt, features such as non-absoluteness, context-sensitivity, and relationality are fundamental to social-relational approaches to moral status. In this way, we see that even in taking the relational turn we need not abandon outright the use of properties in ascribing moral standing. Yet it remains of utmost importance to assure that the properties we may employ as benchmarks for locating moral status are themselves relational.¹⁹

5. Conclusion

To conclude, I highlight one of the key implications of my application of Kantian revisionism to AI and robots, namely the idea that technological entities with which we interact are granted moral status in similar ways as non-human animals and even fellow humans are granted moral status. Indeed, it may be that, to some of us, all such others are afforded a similar degree of moral standing. In other words, we may see similar reasons and similarly forceful demands to respect all others. And this may come across as rather controversial. I suspect that some will disagree with the thought that respect for robots could be on par with respect for fellow human beings. However, here I reiterate that I have focused not necessarily on the sorts of behaviors we might engage in, but rather, on how to justify maintaining respectful relations, and framed as such, it does seem—at least on the framework I outlined—that we might reasonably find a similar justification for respecting robots as we find for respecting non-human animals and even other humans. That being said, a first caveat to the account I have offered here is that some will not accept the revisionist account of Kant. While I find Sensen's analyses provocative yet persuasive, some readers will understandably be too wedded to the standard reading. For them, dignity is and should remain an absolute inner value, which, if we find it, serves to justify our moral treatment of others.

As a second closing caveat, I should note that what I have argued here is that we may find grounds for treating AI and robotic systems with *respect*. But this does not necessarily entail granting them *rights*. I leave it to interested readers to make the connection, or perhaps find ways of obscuring the connection, between respect and rights. While I am somewhat sympathetic to recent critics who argue against technological solutionism and the inevitable place of AI and robotics in our society (cf. Bryson 2018), as far as I can tell, the lack of inevitable rights does not entail an inevitable lack of rights. If we accept my opening observation, namely that we seem to spend a great deal of our time today with interactive technological systems, it seems fair to suppose that at least some of us will develop some sort of relationship with them. Some of us are already seeing some systems as part of our social and moral worlds.²⁰ Perhaps accounting for this movement will require wholesale paradigm shifts in our moral theorizing, as Coeckelbergh suggests. But as I hope to have shown here, the movement toward respect for robots might also be bolstered by moderate revisions of classic ethical theories.

Endnotes:

¹ Schönecker and Schmidt note that some traditionalists maintain a form of moral realism to support value as a metaphysical property (e.g. Wood 1999), while others take a more constructivist approach (e.g. Korsgaard 1996). Nonetheless, what they call the 'standard view' of value, they claim, is also held by scholars such as Paton (1971); Horn, Mieth, and Scarano (2007); and Allison (2011).

² See Sensen (2011) for a summary, and textual examples from political documents, illustrating the contemporary view of dignity as an intrinsic value that grounds claims to respect.

³ For this reason, philosophical accounts of abortion sympathetic to Kantian ethics often employ a supplementary notion, like potentiality (Hare 1989) or capabilities, namely of a woman to determine the course of her life (Dixon and Nussbaum 2012).

⁴ Here, deontologically inclined ethicists can invoke Kant's notion of indirect duties. In short, this is the idea that we can have duties to non-human animals—and perhaps to the environment—since our treatment of them affects our character and thereby our treatment of other humans. See Kant's *Lectures on Ethics*; also O'Neill (1998).

⁵ For more on exclusionary mechanisms toward machines, see Gunkel (2018, pp. 24-39).

⁶ This summary is only a rough sketch of Sensen's detailed reading, translation, and analysis—particularly of Kant's addendum to the Formula of Autonomy in the *Groundwork*. In the interest of proceeding to the applications at stake, I must forego a longer explication and I encourage readers to look directly at Sensen's insightful work.

⁷ Kant's *Lectures on Pedagogy*, 9:488; as quoted in Sensen (2011, p. 81).

⁸ As Sensen argues at length, the source of normativity for Kant is found in the categorical imperative itself (and not in the apparent dignity of others, since dignity is self-referential). Again, for a much fuller demonstration and rich textual evidence, I refer readers to Sensen's analyses.

⁹ This is why many authors work to revise, or loosen, the conditions for agency (e.g. Floridi and Sanders 2004) in order to accommodate AI, or are otherwise content to speak of alternative (less full) conceptions of agency—such as functional agency or 'artificial moral agents' (e.g. Allen and Wallach 2008). I return to these considerations below.

¹⁰ While Nyholm's investigation seeks a more direct argument, the account stops short of supporting respectful treatment for robots. The account herein concludes similarly, except that on the revised Kantian picture, reasons to respect others flow from one's own dignity, rather than a recognition of 'apparent humanity' as seen in Nyholm (2020). Additionally, my account differs in relying on a nuanced understanding of dignity (see previous section), while Nyholm speaks more broadly, for example, of treating robots 'with a certain amount of respect and dignity' (ibid., p. 187).

¹¹ Although Gunkel's (2018) focus is primarily centered on arguments for and against the factual possibility—and the normative stances—concerning robot rights, his discussion applies equally well here to the idea of treating robots with respect (whether or not that entails any sort of right). I will return to this distinction in closing.

¹² For further discussion, see Coeckelbergh (2021). Here Coeckelbergh enumerates and insightfully defends four conditions for indirect moral standing for social robots. His second condition, where a user has developed feelings of attachment and empathy, is most applicable to the present reasoning. I further discuss the four considerations below.

¹³ Notably, at least one quasi-religious movement of this sort—known as *Way of the Future*—has already surfaced and, recently, has disbanded. The movement was indeed said to be founded partly on fear of the 'singularity' and was committed to the 'safe transfer of power between humans and their up-and-coming AI overlord' (Buck 2021). As R. Anthony Buck suggests, in an article for the *European Academy on Religion and Society*, 'chances are it will not be the last religion to worship AI.'

¹⁴ It seems far from clear who *could* be in a position to advise us on such matters. Indeed, the prudential value of worshipping an AI god is surely an area for future research and ripe for interdisciplinary perspectives, hopefully including technology and religious experts, among others.

¹⁵ For encouraging me to clarify these important points, I am grateful to an anonymous reviewer.

¹⁶ For related arguments, see, Danaher's (2020) account of ethical behaviorism, as well as Coeckelbergh's (2021) fourth argument for indirect moral standing, concerning how we should treat others when in doubt of their moral standing. Also, for comparable discussion, see Gunkel (2012, p. 176).

¹⁷ Here I am grateful for comments from an anonymous reviewer.

¹⁸ In fact, in previous writings, I have subscribed precisely to the focus on relations and interactions over and above properties of the other (e.g. Tigard 2021a; 2021b; 2021c).

¹⁹ Since here I have both resisted some of Levinas's thinking—namely the rejection of properties—and at the same time abided by his calling, namely to rethink philosophical methods in ways that do not reduce differences, I follow Coeckelbergh (2021, p. 19) and Gunkel (2012, p. 182) in using Levinas against himself. Also, I should note that my account has retained the use of notions like moral agent and patient and so, I take it, cannot be classified as fully thinking otherwise, on Gunkel's analysis. Still, with its reliance upon an interpretation of dignity that is by its nature relational, I find that the application of Kantian revisionism is a step toward deconstructing a strict demarcation of agents and patients.

²⁰ For several real-life examples of this movement, see Chouwa Liang's insightful *New York Times* documentary: <https://www.nytimes.com/2023/05/23/opinion/ai-chatbot-relationships.html>

References

- Allison, H. E. (2011). *Kant's Groundwork for the Metaphysics of Morals: a commentary*. Oxford University Press.
- Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford UP.
- Breazeal, C. (2003). Toward sociable robots. *Robotics and autonomous systems*, 42(3-4), 167-175.
- Bryson, J. J. (2018). Patency is not a virtue: the design of intelligent systems and systems of ethics. *Ethics and Information Technology*, 20(1), 15-26.
- Buck, A. (2021). The Way of the Future is now a thing of the past. *European Academy on Religion and Society*. <https://europeanacademyofreligionandsociety.com/news/the-way-of-the-future-is-now-a-thing-of-the-past/>
- Coeckelbergh, M. (2010). Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology*, 12(3), 209-221.
- Coeckelbergh, M. (2021). 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(3), 337-360.

- Danaher, J. (2020). Welcoming robots into the moral circle: a defence of ethical behaviourism. *Science and Engineering Ethics*, 26(4), 2023-2049.
- Darling, K. (2016). Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects. In Calo, R., Froomkin, A. M., & Kerr, I. (Eds.) *Robot Law* (pp. 213-232). Edward Elgar Publishing.
- Dixon, R., & C. Nussbaum, M. (2012). Abortion, Dignity, and a Capabilities Approach. In B. Baines, D. Barak-Erez, & T. Kahana (Eds.), *Feminist Constitutionalism: Global Perspectives* (pp. 64-82). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511980442.006
- Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, 42(3-4), 177-190.
- Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines*, 14(3), 349-379.
- Good, I. J. (1965). Speculations concerning the first ultraintelligent machine. *Advances in Computers*, vol. 6, 31-88.
- Gunkel, D. (2018). *Robot rights*. MIT Press.
- Gunkel, D. (2012). *The machine question: Critical perspectives on AI, robots, and ethics*. MIT Press.
- Gunkel, D. (2020). *The Year of Robot Rights*. MIT Press Reader. <https://thereader.mitpress.mit.edu/2020-the-year-of-robot-rights/>
- Hare, R. M. (1989). A Kantian approach to abortion. *Social theory and practice*, 15(1), 1-14.
- Horn, C., Mieth, C., & Scarano, N. (2007). *Immanuel Kant. Grundlegung zur Metaphysik der Sitten*. Suhrkamp: Frankfurt.
- Kant, I. (2008). *Groundwork for the Metaphysics of Morals*. Yale University Press.
- Kurzweil, R. (2005). *The singularity is near: When humans transcend biology*. Penguin.
- Levinas, E. (1969). *Totality and infinity* (A. Lingis, trans.). Pittsburgh: Duquesne University Press.
- Nyholm, S. (2020). *Humans and robots: Ethics, agency, and anthropomorphism*. Rowman & Littlefield Publishers.
- O'Neill, O. (1998). Kant on Duties Regarding Nonrational Nature. *Proceedings of the Aristotelian Society Supplementary Volume*, 72(1), 211-228.
- Paton, H. J. (1971). *The categorical imperative: A study in Kant's moral philosophy*. University of Pennsylvania Press.
- Schönecker, D., & Schmidt, E. E. (2018). Kant's Ground-Thesis. On Dignity and Value in the Groundwork. *The Journal of Value Inquiry*, 52(1), 81-95.
- Singer, P. (1975). *Animal liberation. Towards an end to man's inhumanity to animals*. Granada Publishing Ltd.
- Sensen, Oliver. (2009). Kant's Conception of Human Dignity. *Kant-Studien*, 100(3), 309-331.
- Sensen, O. (2011). Human dignity in historical perspective: The contemporary and traditional paradigms. *European Journal of Political Theory*, 10(1), 71-91.
- Tigard, D. W. (2021a). There is no techno-responsibility gap. *Philosophy & Technology*, 34(3), 589-607.
- Tigard, D. W. (2021b). Artificial moral responsibility: How we can and cannot hold machines responsible. *Cambridge Quarterly of Healthcare Ethics*, 30(3), 435-447.
- Tigard, D. W. (2021c). Artificial agents in natural moral communities: A brief clarification. *Cambridge Quarterly of Healthcare Ethics*, 30(3), 455-458.
- Wada, K., & Shibata, T. (2006). Robot therapy in a care house-its sociopsychological and physiological effects on the residents. In *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006*. (pp. 3966-3971). IEEE.
- Wallach, W., & Allen, C. (2008). *Moral machines: Teaching robots right from wrong*. Oxford University Press.
- Wood, A. W. (1999). *Kant's ethical thought*. Cambridge University Press.

Received: 02/12/2022

Revised: 15/06/2023

Accepted: 24/06/2023