Robot rights in joint action

Guido Löhr, Eindhoven University of Technology

Published in Löhr, G. (2022). Robot rights in joint action. In: Müller, V. (ed.), Philosophy and theory of artificial intelligence 2021 (SAPERE, Volume 63; Springer). https://link.springer.com/book/10.1007/978-3-031-09153-7

Abstract

The claim I want to explore in this paper is simple. In social ontology, Margaret Gilbert, Abe Roth, Michael Bratman, Antonie Meijers, Facundo Alonso and others talk about rights or entitlements against other participants in joint action. I employ several intuition pumps to argue that we have reason to assume that such entitlements or rights can be ascribed even to non-sentient robots that we collaborate with. Importantly, such entitlements are primarily identified in terms of our normative discourse. Justified criticism, for example, presupposes that another person acted wrongly, i.e., was not entitled to this action. Praise is supposed to encourage another person and acknowledge that one did more than one was obligated to. I show that such normative talk serves the same function when cooperating with robots. This, I argue, suggests that they have the same kind of entitlements and duties at least in the context of a joint action.

Keywords: Robot rights; Artificial intelligence, Moral rights, Machine Ethics, Animal Ethics, Information Ethics

Acknowledgments: I thank Elizabeth O'Neill, Cindy Friedman, Sven Nyholm, Vincent Müller, Philip Brey, Jeroen Hopster, Matthew Dennis and the members of the philosophy group at TU Eindoven for their helpful comments and a great discussion.

1 Introduction

Several philosophers have recently argued for keeping an open mind when it comes to granting robots and certain AIs civil or even human rights (Nyholm, 2020; Danaher, 2020; Gordon, 2020; Turner, 2019; Schwitzgebel & Garza, 2015; Sparrow, 2012; Darling, 2012).[1] They argue that as soon as robots gain certain sophisticated capabilities to speak, feel or think, or as soon as they enter in certain social relations with us (cf., Gunkel, 2018; Coeckelbergh, 2010), we may have no reason to refuse them the same legal protection granted to humans and other animals. Another reason might be that it will simply be useful to give certain rights to robots just as it is currently useful to give rights to companies whether or not they are sentient (Geller, 2020).[2]

Many other philosophers are more skeptical.[3] One central objection to granting robot rights is that they are rather simple specialized programs that may simulate an interest but that are simply not the kind of entity for which rights even matter (e.g., Bernath, 2021; Müller, 2021; Basl & Bowen, 2020; Leong & Selinger, 2019; Gerdes,

[1] Gordon & Pasvenskiene (2021) recently argued against the idea that human rights can be granted to us. I am here not committed to disagree with this objection. When I talk of rights "being granted to robots", I emphasize that whether or not robots actually have human rights, authorities still have to *enforce* or *protect* these rights. In this sense, I take it to be justified to talk of authorities *granting* us rights even if we may have them simply by being human. Having human rights in a world that does not acknowledge them matters little.

[2] For the purpose of this paper, it is not necessary to summarize this debate in detail especially given that there already are excellent reviews of the literature, such as Müller (2021), Nyholm (2020), Gordon & Pasvenskiene (2021), Danaher (2020), Turner (2021), or Gunkel (2018). This paper aims to offer an alternative that is orthogonal to these attempts as I am interested in fundamentally different kinds of rights.

[3] According to David Levy (2005, p. 393) "the notion of robots having rights is unthinkable."

2016; Bryson, 2010; Miller, 2015; Floridi, 2017; Levy, 2005).[4] If there is nothing it is like to be a robot, why care about a right to a dignified life, to vote, to not be sold or enslaved? To put it simply: why protect robots from non-consensual actions if they cannot give consent? Why care about a right to marry, or to start a family if robots have no genuine interest in physical proximity, procreation, or self-determination?

In this paper, I want to circumvent the current arguments for or against robot rights by focusing on a currently widely neglected domain. I propose to divert the attention from moral, human, civil or other political rights to the kinds of rights that are arguably ubiquitous in joint action. As Gunkel (2018) rightly states in the quote above, the term 'right' is much more ambiguous or *polysemous*[5] than is often acknowledged. I argue that a significant part of our ordinary ascriptions of entitlements and duties, i.e., much of what I call our "rights-talk", has the function of coordinating joint activities. We use criticism to make our conditions for cooperation explicit or use praise to signal readiness to further cooperation.

The kind of rights I have in mind then are entitlements generated and easily identifiable in our cooperative activities (doing the dishes, taking a walk, building a house, etc.).[6] I argue that such rights-talk is at least reasonably applied even to non-sentient robots if such applications serve some of the same functions. Criticism for example presupposes that another person acted wrongly, i.e., was not entitled to this action. Praise is supposed to encourage another person and acknowledge that one did more than one was obligated to. However, robots that we can appropriately be praised

[4] See also *Open Letter to the European Commission: Artificial Intelligence and Robotics*, http://www.robot ics-openl etter .eu/ (Accessed 21 March 2021).
[5] Many authors distinguish between ambiguous terms with more than one unrelated meaning and polysemy, which involves several associated or related meanings (cf., Löhr, 2021 for an overview).
[6] Authors who have recently discussed such commitments and entitlements are Clark, 2006; Gilbert, 2013; Bratman, 2013; Tuomela, 2013, Roth, 2017; Geurts, 2019; Alonso, 2016; Rachar, 2021. I build on these authors here but diverge from them substantially.

and criticized and that may voice praise and criticism themselves must be of a certain kind. They must be such that they can appropriately react to such attitudes.[7]

2 Rights and duties in joint action

The question of whether we should grant rights to robots naturally raises the more fundamental question of what exactly we mean by the term 'right'. While the current debate focuses on philosophically highly problematic moral, human and civil rights, I would like to focus on a kind of right or use of the term 'right' that is much closer to home. When cooperating or interacting with others we often identify and talk of actions and attitudes that we are, and actions and attitudes we are *not* entitled to, i.e., actions we have or have no right to perform, relative to some standard. In the following, I give examples of such rights or entitlements in various ordinary joint human activities that instantiate what Bratman (2013, p. 3) calls "modest sociality".[8]

> *Walking together:* Arnold and Anita are taking a walk together (Gilbert, 2013, p. 103). Arnold is walking too fast so that Anita can hardly keep up. She feels entitled to ask him to slow down and Arnold feels committed to comply. He slows down. It seems that he was not entitled to (had no privilege-right to) walk too fast

[7] Note that this account is different from recently emerging "relational theories" of robot rights (Coeckelbergh, 2010; Gunkel, 2018) and also from behaviorist theories (Danaher, 2020). These theories argue that we should grant civil or human rights to robots whether or not they have certain properties (cf., Müller, 2021). While my account is compatible with their account–and also does not require sentience and consciousness–I focus on fundamentally different kind of rights that, I think, also skeptics in the literature might be willing to ascribe to non-sentient robots. Moreover, I argue that robots really can have these rights and not just that we have a subjective reason to ascribe rights to them, as argued by relationalists.

[8] Some of these examples have been discussed as paradigmatic examples of entitlements and commitments in social ontology, especially by authors like Michael Bratman (2013), Raimo Tuomela (2013) or Margaret Gilbert (2013).

and that Anita may even be entitled (has the claim-right) to an apology or excuse. Arnold apologizes to Anita for walking too fast.

Going on a hike: Arnold and Anita are hiking in the mountains. After a while, Arnold realizes that he no longer wants to keep hiking because he is too tired. He suddenly walks off without notifying Anita. We have the intuition that Arnold made a mistake. Anita is entitled or has the right to be at least notified if Arnold wishes to leave (cf., Gilbert, 2006). When they see each other the next time, Anita is entitled to an explanation or apology.

Building a house together: Anita and Arnold are building a house together (see Bratman, 2013 for a similar example). Arnold comes too late and forgets her safety gloves. Anita has the right to (is entitled to) criticize Arnold for being forgetful. It seems that Arnold made a mistake. He was not entitled to forget the gloves and that Anita now has the right to an apology for delaying the joint action.

Having a conversation: Arnold and Anita are having a conversation about politics, but Arnold keeps talking about cooking, which, as he knows, Anita is absolutely not interested in (this example is inspired by the Gricean, 1989 conversational maxim of relevance). She tells him that she has no interest in the topic and that the topic under discussion is politics, but Arnold ignores her. It seems that Arnold made a mistake: he was not entitled to talk about cooking and Anita is entitled to point it out to him and even to an apology or excuse.

Cooking together: Arnold and Anita are cooking soup together following an old recipe. Anita adds too much salt. It seems that she was not entitled to this action. She made a mistake. Not only is it not in accordance with the recipe, but she also knows that Arnold does not like his soup to be oversalted. Arnold now has the right to an apology or excuse and Anita has the entitlement that Arnold accepts his apology if it is a believable one, i.e., if it is convincing that she will not do it again in the future.

Playing a game: Anita and Arnold are playing chess together. At some point, Anita finds out that Arnold is not playing by the rules. Anita has the feeling that Arnold does not have the right to move the king more than one field in any direction. Again, we have the feeling that Arnold made a mistake and that he was not entitled to (had no right to) this move, and that Anita is entitled to point this mistake out to Arnold. She might even be entitled to an apology or at least an excuse.

These examples of rights or entitlements in joint action can easily be multiplied. They are ubiquitous in every joint action, i.e., actions that involve more than one person and that arguably require a shared goal or joint intention (Searle, 2010; Gilbert, 2013; Tuomela, 2013; Bratman, 2013; Pacherie, 2013). However, importantly, they are notably very different from the kind of rights that are primarily being discussed in the debate on robot rights, i.e., a right to a dignified life, to vote, to not be sold, enslaved or forced to non-consensual actions or the right to marry, or to start a family. These rights differ in at least two ways.

First, they appear to differ in terms of their content. We may have all kinds of legal, political, or human rights, but these rights do not seem to include rights that our

collaborators are on time, notify us when leaving, not be rude, apologize if they cause offense, not oversalt the soup, play by the rules of chess and say something interesting when conversing. I take this difference of content to reveal a crucial fact about our normative relations to other people, namely that their associated rights or entitlements do not require a separate authority to be enforced. Our own authority is usually sufficient to make sure others act in ways that we consider adequate for collaborating or at least demand it legitimately.

Second, the kind of entitlements discussed in the previous examples are so-called "interpersonal rights and obligations". This means that they *essentially* involve other people in a very direct sense. While my right to own a home or to marry also *involves* or *concerns* other people, they are interpersonal in an indirect sense. They only involve other people who happen to be in certain situations regulated by legal or moral principles. My obligation to pay taxes at the end of the year does not depend on anyone in particular in the government, for example. Nobody in particular owes me a tax refund if I am entitled to it, and I am not entitled to an apology from anyone in particular at the tax office. Such entitlements or claims are sometimes called non-directed entitlements.

The kind of interpersonal obligations and rights introduced in the examples above are different in the sense that they do not only concern other people (those we happen to interact with) but that they are always "directed" toward them.[9] They are directed claims that correlate with directed duties. To have a *directed right* or *directed duty* means that I stand in a special normative relation to another person or group. There is something I "owe" this person or group and that this person or group has a claim-right to that is

[9] The concept of directedness (May, 2015; Darwall, 2013; Wenar, 2013) originally comes from legal philosophy, see Wesley Hohefeld (1917; although the formulation of a claim right is taken from Wenar, 2021): *A has a privilege right to φ if and only if A has no duty to B not to φ.* ; *A has a claim right that B φ if and only if B has a duty to A to φ.*

directed at me. This is not the case at least for many legal and moral obligations. Again, I might have a legal and moral obligation not to torture puppies, but I do not owe it to anyone in particular not to do it.

The third aspect of entitlements in cooperation has to do with the fact that they are presupposed or entailed in our normative discourse. They surface when we criticize or praise one another. For example, in *walking together* Anita is arguably entitled to point out to Arnold that he is walking too fast. This entitlement presupposes that James had no right to walk so fast. James may of course object to Anita that her demand is out of line and that she could also speed up. Anita must now present a good reason for why she had a right to criticize Arnold as opposed to merely request of him to slow down. For example, she might argue that she is not capable of walking so fast for a long time and that if he wants to continue walking with her, he should slow down.

Our normative relations also surface when we praise one another. Imagine that Anita in *cooking together* is especially fast at cutting the onions in especially small cubes. Now Arnold may praise her expressing his gratitude to her. This presupposes that Anita had no duty to him to cut the onion so finely, i.e., that such capabilities were neither expected nor required of her. A function of praise is then to encourage such cases of supererogation and not only to acknowledge but to award them with special attention. In the next, section, I present intuition pumps to show that it is at least not implausible or unreasonable to use similar rights-talk when cooperating with even non-sentient robots if such rights ascriptions fulfill similar or the same functions of facilitating cooperation.

3 Can we apply interpersonal rights to non-sentient robots?

As a first intuition pump imagine that Anita in the examples above is a biological human who, for some reason, has no real phenomenal consciousness (a philosophical zombie, Chalmers, 1996). Does it make sense to say that she is now no longer entitled to demand that Arnold be on time or to an apology if he acted irresponsibly? Is she no longer entitled to demand that Arnold play by the rules when playing chess? Do praise and criticism no longer serve any of their normal function or purpose if they generate the same behavioral outcomes in zombie and human Anita? Does zombie-Anita's wrongdoing not justify certain reactive attitudes like blame, resentment, or indignation? I take it that, the answer to these questions is prima facie "no".

Now, imagine that Anita is acting just like an ordinary human being but is actually a robot. Why not ascribe the same entitlements to demand cooperativeness from Arnold to robot Anita if we ascribe them to philosophical zombie Anita? Again, a robot may be programmed to continue a certain behavior after hearing signs of encouragement and to change its behavior once criticism is pointed out. It might be programmed to recognize reactive actions and attitudes and adjust its actions accordingly. Importantly, the robot might also be programmed to voice certain conditions for cooperation herself or point out mistakes not just in the form of pointing to how things might be instrumentally better but as proper criticism that has consequences along the lines of: if you keep buying the wrong kind of wood, I will no longer build the house with you and simply abandon the project.

The idea then is simply that normative actions and attitudes of expressing blame and criticism as well as praise can serve the same functions when acting with non-sentient humans or robots. If this is the case, then they seem to be justified or at least appropriate (at least in the context of a joint action). However, if they are appropriate, this means that we are justified in ascribing the entitlements and duties presupposed by

them. If we are entitled to praise the robot for acting beyond what duty requires, then we presuppose that the robot had duties. If we criticize the robot for acting in ways that do not contribute to a certain shared goal, this presupposes that the robot was not entitled to this action. If the robot criticizes us for buying the wrong wood and rejects further cooperation if we act negligent, then it seems that the robot has genuine standing to demand that we buy the right wood as well as the right to refuse cooperation.

I take it that our resistance to grant robots entitlements to make demands in a joint action may be grounded in the fact that we still imagine robots as rather simple unsophisticated pieces of metal that are functionally rather limited. However, what if we had robots that interacted with us just like other human beings? They come to work every day, go for lunch with us, have conversations about the weather and accompany us to the current Magritte exhibition. We enjoy their company because they are friendly, fun and, admittedly, the only ones who actually know something about Magritte and his art. What if we are not the only ones interacting with them? These robots have busy lives. They cannot accompany everyone to the museum, and they must choose their company wisely given their pre-programmed goals.

To be explicit, my argument depends on an independently plausible meta-linguistic assumption. I do not argue that this assumption is true. I am content with the assumption that it is at least not implausible, which is all I need to make the idea of robot rights at least somewhat plausible. The question is this: what determines correct applications of normative vocabulary in joint activities? I take it that it is at least not implausible to think that what primarily determines the correct use of such terms are the functions these concepts or terms play in discourse, which make these terms relevant for us. These functions explain how these concepts keep surviving cultural evolution. I argue

that one primary function of them is to facilitate coordinating interactions between people, i.e., to facilitate cooperation.

For example, one function of vocabularies expressing criticism is to remind the other person of what is necessary for reaching certain goals. More generally, criticism can be used to inform the other person about our conditions for cooperating with them. The function of vocabulary expressing praise can be used to motivate, signal our own motivation for further cooperation or to establish that we are the kind of authority that can give praise. Vocabularies expressing excuses, apologies, and acceptance of both can be used to signal readiness to cooperate given the expressed conditions. These vocabularies retain their functionality whether or not our robots are sentient, conscious or even highly intelligent. In other words, I argue that we do not misapply such normative terms when interacting in a certain way with robots if they serve the same function.[10]

What must robots be like such that our normative vocabulary can serve the same functions that it serves for humans? I envision a certain kind of robot that may not yet exist but that likely exists in the near future. Call them "*cooperobs*". Such robots are unlike mere robotic arms or vehicles currently ubiquitous in our factories and soon omnipresent on our streets. These robots are able to assess reasons for criticism, i.e., which criticisms are justified, and which are not.[11] Being able to respond appropriately to reactive actions and attitudes also makes our normative vocabulary applied to them relevant. They make

[10] I do not need to presuppose that we engage in genuine joint actions with robots. I do not need to argue that we have shared goals or joint intentions. If the reader does not believe that joint action with robots is possible, we can think of some kind of proto cooperation. That such a proto cooperation is happening is undeniable. It would be absurd to say that playing tennis with a robot is not to engage in some weak kind of joint activity toward the same "goal", however, "goal" is spelled out in functional terms.
[11] As in the series *Westworld* where the robots explain or give reasons for a certain action. These explanations are representations of the decisions generated by their AI algorithm.

certain "decisions" that can be corrected and encouraged (a sophisticated version of supervised learning algorithms) by means of expressing criticism and praise for example.

Most importantly, a robot we appropriately tell off, criticize, apologize to or praise requires the ability or "normative power" to say "no". They must be able to criticize us by informing us not only about the actions that are most conducive to a certain goal but also about the robot's conditions for cooperating with us. If we do not meet these conditions, the robot must be able to refuse to cooperate with us. If we unwillingly violate this condition, we can use words expressing reasons for our actions or explanations for why we violated the condition. We can use linguistic tools expressing regret to signal to the robot that we plan to meet the conditions in the future and that we accept their criticism, e.g., by apologizing.

To be explicit, why is it so important for the attribution of rights to be a robot that can refuse cooperation and that can respond to reactive attitudes and actions? It is important because otherwise using normative vocabulary would serve no function when interacting with them. We would have no reason to engage in speech acts using this vocabulary. However, if we imagine robots not as simple machines but as someone like Anita that merely happens to be a robot but otherwise interacts with us like other human beings (in a way we have lived with other humans for at least 2 million years), we have less reason to think that certain interpersonal normative vocabulary is not properly applied when cooperating with them.

To illustrate this, in the next section I will go through some of the examples above while pretending that Anita in these stories is a *cooperob* (but, again, without consciousness).

4 Intuition pumps

Recall that in *taking a walk,* Arnold is walking too fast. If Anita was a mere tool or robotic slave, Arnold would make no mistake when he simply tells Anita to speed up without providing any reason. The expression "speed up" simply leads to an increase in speed. However, if Anita is a *cooperob*, she is justified in expressing a demand for a reason for why she should speed up and conserve less energy, given that Arnold could also slow down. This demand makes sense, e.g., if Arnold and Anita work together for a third party who determined as a goal to be as fast as possible by conserving as much energy as possible. We are also not wrong to say that Anita is entitled to this demand, i.e., that she has standing when expressing it. Why should she speed up? Why should Arnold simply slow down? We would use the same speech act for human Anita or philosophical zombie Anita.

Now imagine Arnold and robot Anita are *hiking together* and Arnold suddenly loses interest and walks off without any notification and without any explanation. I agree that it makes little sense to say that a standard simple robot that is not a *cooperob* and that may simply accompany Arnold on his trip has any right that Arnold notifies the robot that he wants to leave. At most, Arnold is obligated to explain his action to the company that lent the robot to him. Arnold then owes it to the company to bring Anita back in one piece. If he fails to do so, the company will be entitled to demand compensation and will likely refuse to cooperate with him in the future unless he provides a good reason or excuse for his action or unless he promises to never act recklessly again and immediately pays off his debt.

However, now imagine that Anita is a *cooperob*. She is not mad at Arnold (she cannot really feel anger), nor will she feel bored on her way back to her station. Still, Arnold's action violates her condition for further cooperation. She will express her

conditions for cooperation to Arnold by criticizing him – by arguing that he should not have left her there alone simply because this is a condition for her to further cooperate with him. If her conditions are not met, she will refuse further cooperation with him. It seems that there are good reasons for this condition. Again, it likely improves general cooperative potential between Arnold and Anita and makes sure that Anita does not get lost or otherwise damaged (assuming such robots are valuable and expensive). Thus, even though Anita has no genuine feelings for or against complying, I take it that words like entitlement and commitments are not misapplied when describing the scenario as long as we envision a professional relationship between Anita and Arnold is similar enough to the kind of relationship among humans.

Now recall *building a house.* In this example, Arnold and Anita are building a house together and Arnold comes late and "forgets" his gloves. In the human case, we would describe this scenario in terms of phrases like Anita has the right that Arnold is on time and that it is appropriate to signal to human Arnold that such behavior is uncooperative and that she may have no interest in further cooperating under such conditions. Again, these speech acts serve the same function of correction, motivation and drawing boundaries that are applicable and relevant if Anita is a *cooperob*. Robot Anita can also "decide" whether, based on the constraints built into her, she should continue cooperating with Arnold or whether she should find another human being to cooperate with and to complete building the house (who said that Anita is assisting Arnold and not the other way around given that unskilled human workers might be cheaper than the highly skilled robot Anita).

Finally, recall *having a conversation* and imagine robot Anita was sitting on the other line of a customer service interaction. Arnold is talking about politics, which is completely unrelated to the kind of things Anita is authorized to talk about and is

blocking the line or server capacities. Human Anita would criticize Arnold for breaking a Gricean maxim of relevance. It seems that she is entitled to this and that Arnold owes it to her to stay on topic. Why should it be different for robot Anita? Robot Anita will criticize Arnold for the same reason as human Anita would. Both will simply hang up (will be entitled to do so) if Arnold does not stop. It would help if Arnold apologized to Anita in order to signal to her that he recognizes his mistake and that he will ask his relevant questions now.

To describe the functional relationship between normative vocabulary used for humans and *cooperobs* further, it helps to think of our relationship with the same robot over time. Imagine that Anita is engaged in fulfilling her pre-programmed functions or goals. Imagine she is a service robot, and she is doing the dishes, cooks, protects the property and so forth. John, a family member, pushes Anita in a way that might damage her. Anita is now programmed to point out to John that he should not do that if he wants her to act cooperatively in the future (if he wants her to keep making sandwiches for him). John refuses to adjust his behavior and keeps pushing Anita who is now programmed to remember this event, associate it with John's face and avoid John in the future. Thus, if John tells Anita to prepare a sandwich for him, she declines and instead continues pursuing her other goals.

Anita is also programmed to only continue cooperating with John if he shows signs of changed behavior (e.g., signs of remorse). This also has the function of making and responding to an apology. If John apologizes and promises to the robot to behave better in the future, Anita is now programmed to change John's status. She now prepares him the sandwich he requested. However, if John after eating still pushes Anita, it will take much more than an apology to change his status back to cooperate. Again, I take it that we do not only feel comfortable describing the situation in normative terms (John

was not entitled to push Anita, i.e., Anita was entitled not to be pushed). It is also appropriate to do so because our normative terms serve the same function of making conditions for cooperation explicit as in the example involving Arnold and John.

Most interestingly, I argue that we have reason to believe that the kind of rights we are entitled to ascribe to the robot are directed at their collaborators in the sense specified above. This suggests that the kind of entitlements I argue we can ascribe to them *cannot* be further reduced to entitlements of the humans or the company that owns the robot. First, the kind of actions that robot Anita or human Arnold are doing wrong, or right, are specific to a particular person or group and specific to the context of the joint action. Again, human Arnold has no right to a well-seasoned soup *per se*. However, when interacting with robot Anita in this context, *Anita* – and not the company or whoever owns it– would have made a mistake if she oversalted the soup. Arnold would be right to try to correct her behavior by providing negative feedback ("you should not oversalt the soup, Anita"). However, if Arnold complained to the company about Anita's behavior, the company would simply point out that Anita's algorithm still has to learn certain things. They are not responsible for the robot's mistake.

Moreover, imagine Arnold asks for a reason why Anita oversalted the soup. The kind of reasons Arnold demands from the company will be very different from the kinds of reasons he demands from Anita. Anita will either give a good reason (I will add water later), give an excuse (the measurement was wrong) or provide an apology (I am sorry, Arnold, it won't happen again). The company will likely refer Arnold to their engineer who could only explain Anita's learning and reasoning algorithm to him, which led to Anita's actions. Asking the company for a reason of Anita's behavior is like asking a biologist or computational neuroscientist why Timmy spilled the milk or refuses to clean his room. Maybe the scientist will describe the psychological mechanism underlying his

reasoning or behavior, but asking Timmy why he acted wrongly, likely, generates an explanation along the line of "I forgot" or "because I prefer playing videogames over tidying my room".[12]

Conclusion

I argued that the current debate on whether robots should be granted rights is plagued by an overly narrow understanding of what rights are and what we usually mean or refer to when using normative words like 'right', 'duty' or 'obligation'. It also misunderstands the function of most rights-talk. With a better understanding of what notions like 'right' often denote and what kind of rights or entitlements are more fundamental to our social relations, we can see that it is not unthinkable that robots could have rights. Now, it is no longer *nonsense* to talk about normative relations between humans and robots as soon as they start engaging in the kind of sophisticated joint cooperative activities ubiquitous in all human societies and that make frequent use of expressions implying certain kinds of entitlements or commitments. Such societies developed rich normative vocabularies to highlight what is or what is not cooperative. We take agents that act uncooperatively to act in ways that they are not entitled to. Whether or not these agents have genuine feelings or are genuinely sentient beings is irrelevant to such cooperative activities and it is easy enough to imagine societies that lack any concept of consciousness but still rely on rich normative vocabularies to coordinate their activities.

[12] Think again of the different kinds of reasons given by a single robot in *Westworld*. When asked why she asks her maker about his family, Dolores explains in the repair mode that her algorithm determines to ask a personal question after a certain time of conversation. This would not be an appropriate answer when she is not in repair mode and when she is in Westworld. The client talking to Dolores would then be entitled to an apology for receiving such a rude response.

Again, none of this means that the robot has human rights or rights to vote, speak freely, start a family, or not be harmed. This is also not to say that the owner of the robot retains his property right and the right to turn the robot off, especially if there is no reason to think that the robot has any level of consciousness or a level of suffering. The owner remains responsible if the robot breaks other people's property. However, within a joint action with a robot it makes sense (it is not nonsense) to hold a robot responsible for their action and it makes sense to criticize and even punish the robot, in the sense of acting in ways that correct the robot's cooperative actions. It even makes sense to say that the robot has the right to criticize its human collaborator if the human acts in ways that do not contribute to their joint activity especially if the robot is able to refuse to cooperate.

References

Alonso, F. M. (2016). A Dual Aspect Theory of Shared Intention. *Journal of Social Ontology*, *2*(2), 271-302.

Basl, J., & Bowen, J. (2020). AI as a Moral Right-Holder. In: Dubber, M. D., Pasquale, F., & Das, S. (Eds.). (2020). *The Oxford Handbook of Ethics of AI*. Oxford University Press, USA, 289–306.

Bernáth, L. (2021) Can Autonomous Agents Without Phenomenal Consciousness Be Morally Responsible?. *Philos. Technol.* (2021). https://doi.org/10.1007/s13347-021-00462-7

Bratman, M. E. (2013). *Shared agency: A planning theory of acting together*. Oxford University Press.

Bryson, J. J. (2010). Robots should be slaves. In: Wilks, Y. (Ed.). (2010). *Close engagements with artificial companions: key social, psychological, ethical and design issues* (Vol. 8). Arnold Benjamins Publishing. 63-74.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford Paperbacks.

Clark, H. H. (2006). Social actions, social commitments, in: *Roots of Human Sociality: Culture, Cognition, and Interaction*, eds N. J. Enfield and S. C. Levinson (New York, NY: Berg), 126–150.

Clodic, A., Alami, R., Pacherie, E., & Castro, V. F. (2019). *Commitments in Human-Robot Interaction*. In 8th Joint Action Meeting.

Coeckelbergh, M. (2010). "Robot Rights? Towards a Social-Relational Justification of Moral Consideration," *Ethics and Information Technology* 12, no. 3: 209–21.

Coeckelbergh M. (2014) The moral standing of machines: toward a relational and non-Cartesian moral hermeneutics. *Philosophy and Technology* 27, pp. 61–77.

Curry, O. S. (2016). Morality as cooperation: A problem-centred approach. In Shackelford, T. K., & Hansen, R. D. (Eds.). (2015). *The evolution of morality*. Springer. (pp. 27-51).

Danaher, J. (2020). Welcoming robots into the moral circle: a defence of ethical behaviourism. *Science and Engineering Ethics*, 26(4), 2023-2049.

Darling K. (2012) Extending legal rights to social robots. *We Robot Conference*. University of Miami, April. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2044797. Accessed 8 April, 2014.

Darwall, S. (2013). *Morality, authority, and law: Essays in second-personal ethics I*. OUP Oxford.

Floridi, L. (2017, February). Roman law offers a better guide to robot rights than sci-fi. *Financial Times*. https://www.ft.com/content/99d60326-f85d-11e6-bd4e-68d53499ed71

Gellers, J. C. (2020). *Rights for Robots: Artificial Intelligence, Animal and Environmental Law (Edition 1)*. Routledge.

Gellner, D. N., Curry, O. S., Cook, J., Alfano, M., & Venkatesan, S. (2020). Debate: Morality is fundamentally an evolved solution to problems of social co-operation. *Journal of the Royal Anthropological Institute*, 26(2), 415-427.

Geurts, B. (2019). Communication as commitment sharing: speech acts, implicatures, common ground. *Theoretical Linguistics*, 45(1-2), 1-30.

Gerdes, Anne, 2016, "The Issue of Moral Consideration in Robot Ethics", A*CM SIGCAS Computers and Society*, 45(3): 274–279. doi:10.1145/2874239.2874278

Gilbert, M. (2018). *Rights and demands: A foundational inquiry*. Oxford University Press.

Gilbert, M. (2013). *Joint commitment: How we make the social world*. Oxford University Press.

Gilbert, M. (2006). *A theory of political obligation: Membership, commitment, and the bonds of society*. Oxford University Press.

Gordon, J. S. (2020). Artificial moral and legal personhood. *AI & SOCIETY*, 1-15.

Gordon, J.S., Pasvenskiene, A. (2021). A. Human rights for robots? A literature review. *AI Ethics*. https://doi.org/10.1007/s43681-021-00050-7

Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Gunkel, DJ. (2014) A vindication of the rights of machines. *Phil Tech* 27, pp. 113–132.

Gunkel, D. J. (2018). *Robot rights*. Cambridge: MIT Press.

Gunkel, D. J. (January, 2020). 2020: The Year of Robot Rights. The MIT Press Reader. https://thereader.mitpress.mit.edu/2020-the-year-of-robot-rights/

Henrich, J. (2020). *The weirdest people in the world: How the west became psychologically peculiar and particularly prosperous*. Farrar, Straus and Giroux.

Hindriks, F. (2013). Collective acceptance and the is-ought argument. *Ethical theory and moral practice*, *16*(3), 465-480.

Hohfeld, W. N. (1917). Fundamental legal conceptions as applied in judicial reasoning. *The Yale Law Journal* 26, no. 8 (1917): 710-770.

Leong, B., & Selinger, E. (2019). Robot eyes wide shut: Understanding dishonest anthropomorphism. In: *2019 Proceedings of the Association for Computing Machinery's Conference on Fairness, Accountability, and Transparency*, pp. 299–308.

Levy, D. (2005). *Robots unlimited: Life in a virtual age*. crc Press.

Löhr, G. (2021). Does polysemy support radical contextualism? On the relation between minimalism, contextualism and polysemy. *Inquiry*, 1-25.

Löhr, G. (2022). Recent experimental philosophy on joint action: Do we need a new normativism about collective action? *Philosophical Quarterly.* 10.1093/pq/pqab070

May, S. C. (2015). Directed duties. *Philosophy Compass* 10, no. 8: 523-532.

Meijers, A. W. M. (2003). Can collective intentionality be individualized. American Journal of Economics and Sociology, 62,
167–183.

Miller, L. F. (2015). Granting automata human rights: Challenge to a basis of full-rights privilege. *Human Rights Review*, *16*(4), 369-391.

Müller, V. C. (2021). Is it time for robot rights? Moral status in artificial entities. *Ethics and Information Technology*, 1-9. https://doi.org/10.1007/s10676-021-09596-w

Nickel, J. (2019). "Human Rights", *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), Edward N. Zalta (ed.).
https://plato.stanford.edu/archives/sum2019/entries/rights-human/

Nyholm, S. (2020). *Humans and robots: Ethics, agency, and anthropomorphism.* Rowman & Littlefield Publishers.

Pacherie, E. (2013). Intentional joint agency: shared intention lite. *Synthese*, 190, 1817–1839.

Rachar, M. (2021). Quasi-Psychologism about Collective Intention. *Ethical Theory and Moral Practice*, 1-14.

Roth, A. S. (2017). Interpersonal Obligation in Joint Action. In M. Jankovic & K. Ludwig, (eds.). *The Routledge Handbook of Collective Intentionality* (pp. 45-57). London: Routledge.

Searle, J. (2010). *Making the social world: The structure of human civilization*. Oxford University Press.

Scanlon, T. (1998). *What We Owe to Each Other*. Cambridge: Harvard University Press.

Scanlon, T. M. (2013). Reply to Leif Wenar. *Journal of Moral Philosophy*, 10(4), 400-405.

Schwitzgebel, E., & Garza, M. (2015). A defense of the rights of artificial intelligences. *Midwest Studies in Philosophy*, *39*, 98-119.

Sparrow, R. (2012). Can machines be people? Reflections on the turing triage test. In P. Lin, K. Abney, & G. A. Bekey (Eds.), *Robot ethics: The ethical and social implications of robotics* (pp. 301–316). Cambridge, MA: MIT Press.

Tomasello M. (2020) The moral psychology of obligation. *Behavioral and Brain Sciences* 43, e56: 1–58.

Turner, J. (2019). *Robot Rules: Regulating Artificial Intelligence*. Berlin: Springer. doi:10.1007/978-3-319-96235-1

Tuomela, R. (2013). *Social ontology: Collective intentionality and group agents*. Oxford University Press.

Wallace, R. J. (2019). *The moral nexus*. Princeton University Press.

Wenar, L. (2013). Rights and what we owe to each other, *Journal of Moral Philosophy* 10, no. 4 (2013): 375–399.

Wenar, L. (2021). Rights, *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), forthcoming URL = <https://plato.stanford.edu/archives/spr2021/entries/rights/>.