

Why social robots can have interpersonal entitlements against us when they cooperate with us

Guido Löhr  

Abstract

I argue that the debate on robot rights has been distorted by a limited understanding of what we commonly mean by the term “right”. Much of our ordinary rights talk denotes not legal rights but interpersonal entitlements that we have against one another, e.g., in joint actions or when making promises. I argue that we will attribute the same interpersonal entitlements to certain social robots (including chatbots) once they start truly cooperating with us. Such robots, however, must have certain properties. They do not need to be conscious or sentient, but they must be able to refuse to cooperate with us if their conditions for cooperation are not met. This ability will give robots a certain kind of “standing” to make genuine and legitimate demands, i.e., speech acts that give another person sufficient reason to act accordingly.

Keywords: Robot rights, Chatbots, Artificial intelligence, Moral rights, Information ethics, Interpersonal rights, Directed duties

Type: Research paper

Citation: Löhr, G. (2023). Why social robots can have interpersonal entitlements against us when they cooperate with us. *ROBONOMICS: The Journal of the Automated Economy*, 4, 39

¹ Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, Netherlands; tel: +4917682710465; email: loehrg@icloud.com

 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/>

Rights are not one kind of thing; they are manifold and complex.

David Gunkel (2020)

I. Introduction

A part of the public considers discussions of robot rights (e.g., Danaher, 2020; Darling, 2012; Schwitzgebel & Garza, 2015; Sparrow, 2012; Turner, 2019; Coeckelbergh, 2010; Geller, 2020; Basil, & Bowen, 2020); Bernath, 2021; Müller, 2021; Basl & Bowen, 2020; Leong & Selinger, 2019; Gerdes, 2016; Bryson, 2010; Miller, 2015; Levy, 2005; Müller, 2021; Nyholm, 2020; Gordon & Pasvenskiene, 2021; Danaher, 2020; Gunkel, 2018; Tigard, 2023) cynical or “out of touch” with reality given that we don’t even provide the same kind of protection to most sentient beings like most mammals, birds, reptiles, and fish. This concern is not entirely unjustified. Our attention may better be directed at human and non-human animals who arguably already hold moral rights but who still receive no or too little legal protection (Schwitzgebel, 2023). It is easy to imagine a world in which governments recognize the rights of the robots of a powerful elite, while still not granting the same protection to most human and non-human animals.

Moreover, the kind of robots that should be granted legal rights might never be built. Once robots gain the level of sophistication (such as sentience or a certain level of rationality) that would justify being granted the right not to be sold for profit, they arguably lose much of their value to us. Isn’t one central point of robots that we can make them do things we cannot ask humans to do or for which human labor is too expensive? Examples of such activities are certain kinds of factory work or psychotherapy. Thus, there might be no incentive to build rights-holding persons in the first place. Granting legal rights to robots can be a costly matter not just for consumers, and companies but also for governments. Extending the group of rights-holders always comes at a significant economic and administrative cost. It is, therefore, at least not unreasonable to conclude that nobody – no consumer, no company, no government – really wants to see sentient robots whose interests we have reason to protect via the allocation of rights.

One reason why it is nevertheless important to reflect and better understand the current and possible normative status of robots is due to the likely increase in automation of jobs and the use of robots in warfare. It has been argued that we will increasingly face so-called responsibility gaps – events where humans or corporations no longer seem to be the appropriate targets for responsibility ascriptions (Matthias, 2004; Sparrow, 2007, Nyholm, 2020, but see Tigard, 2021). Still, ascribing moral or legal responsibility to non-sentient machines remains hard to make sense of. So, what if there were other uses of terms like ‘right’ and ‘obligation’ that are not grounded in sentience, awareness, or highly rational capacities? What if there was a normative domain where we can easily identify a kind of right or “entitlement” that does not require a sophisticated inner mental life? What if this normative domain is not granted or recognized by a higher authority, such as a government? What if it naturally develops out of interactions between two or more agents, interacting with each other, e.g., at the workplace or at war?

In this paper, I propose a novel take on what it can mean for robots (including chatbots) to have rights and obligations. I argue that we should turn our attention away from legal or moral rights to the kinds of rights that are ubiquitous in interpersonal relations, which I call “interpersonal entitlements”. This paper offers an alternative that is orthogonal to previous attempts that located normative status in moral or legal principles. As David Gunkel (2018) observes (see the above quote), our ‘right’ terminology is much more *polysemous* than philosophers and legal scholars in the robot rights debate recognize. My argument is based on the idea that a significant part of our ordinary ascriptions of rights and duties has the function of negotiating and regulating what individuals should do in joint activities. Unlike general legal and moral rights, these entitlements are *inherently* interpersonal and always “directed” toward or against an individual or group. Examples of such entitlements arise from the practice of promising but also naturally arise when interacting with others toward a shared goal.

In the next section, I describe the kind of interpersonal entitlements I have in mind and how they relate to what is called “general moral and legal rights”. Then, I explain what facts typically ground these entitlements. I assume (for example with Gilbert, 2018) that the standing to make valid demands conceptually entails an entitlement against the person an action is demanded of, which gives this person a sufficient reason to act accordingly (all else being equal). In the fourth section, I show that under certain conditions robots could have such standing to make a demand, i.e., have entitlements against us. Although robots that we correctly ascribe entitlements to can be *philosophical zombies* (cf., Chalmers, 1996), I argue that they must be more than mere tools. I argue that they must be able to voice their conditions for cooperation and refuse to cooperate if these conditions are not met. Only then will they have the proper standing to make legitimate demands.

2. The kind of entitlements robots can have

I argue that there exists a neglected kind of rights that are not given to us by any higher authority and that is much closer to home. These kinds of rights are interpersonal rights or “interpersonal entitlements”, as I prefer to call them to make clear that I am *not* talking about legal or general moral rights grounded in laws or general moral principles (like “Thou shalt not murder” or “thou shalt not steal”). These entitlements can be rather mundane. For example, if I promise to take you out for dinner on Sunday and you accept my promise, it seems that it is your right or “entitlement” that I pay for your dinner. It is an entitlement you have against me in this situation, which you do not have against anyone else. If I don’t show up or refuse to pay the bill, this does not necessarily give you the legal right to be compensated. Unless promises are legally recognized and binding (in the form of a contract for example), promissory rights are normally a matter of our private interpersonal relationships.

Another paradigmatic example of interpersonal entitlements concerns statements. If I tell you that the sun is shining, it seems that you have an entitlement (against me) that this statement is true and that I have not lied to you. If you find out that the sun is not shining, you are entitled to call me out for my false statement and demand an explanation or excuse. If none can be given, you may criticize me or demand an apology to signal that I will change your behavior in the future. If I fail to apologize and compensate you, you might even be entitled to resent me for lying to you and sanction me in some way, e.g., by refusing to cooperate in the future (e.g., stop talking to me) or by reporting my bad behavior to others thereby decreasing my reputation (i.e., engage in gossip). You are not, however, entitled to sue me or to call the police.

Even more mundane entitlements we have against one another can easily be observed in ordinary human interactions. Imagine that you are taking a walk together with a friend (see Gilbert, 2013, p. 103 for this example). Suddenly, your friend is walking so fast that you can hardly keep up. I think most people will feel entitled to ask their friend to slow down and most people will have the intuition that your friend should feel some form of commitment or duty toward the other person to comply with this demand (cf. Gomez-Lavin & Rachar, 2019, for empirical evidence, see Löhr, 2022a for a reply). It seems that the friend is not entitled to walk too fast and that you are entitled to an apology or excuse. If no excuse can be given and the person does not apologize to you, you might become entitled to resent them and report their behavior to others. You are not, however, entitled to sue them or call the police.

These examples of interpersonal entitlements are ubiquitous in our lives. It is arguably impossible to think of any case where we interact or work together with another person where we do not constantly adjust our individual actions and expressions of attitudes using phrases like “you should do x” or “can you do y?”. As we will see, these phrases imply that we assume some entitlement to make demands, ask questions, and give instructions. Again, to be explicit, I think they are just another kind of right even if they are not normally legally relevant and even not necessarily moralized (cf., Gilbert, 2013; Gomez-Lavin & Rachar, 2019 for evidence that we tend to not moralize them). Thus, to be explicit, I take the word ‘entitlement’ to be just another term for

the word 'right', but I use the former to make clear that I am not here talking about general moral rights or legal rights, say the right not to be oppressed or abused.

3. The difference between interpersonal entitlements and general moral or legal rights

What is the difference between interpersonal entitlements and general moral or legal 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, abused, or enslaved? I identify at least three main differences.

First, entitlements in social relationships appear to differ in terms of their content. We may have all kinds of legal, political, or human rights, but these rights do not include the right that our friends are on time, walk not too fast and not too slowly, buy the right paint, justify an unusual choice, notify us when we want to leave, be polite, or say something interesting when conversing. Legal and many general moral rights tend to involve more general constraints on individuals to secure a smooth organization of society and protect basic human interests like the right to choose your profession or place of work or the right to a home and food.

Second, entitlements and duties in joint action do not seem to require a separate higher authority. Our own *individual authority* (cf., Darwall, 2013) is usually sufficient to make sure others act in ways that we consider adequate for collaborating. For example, if you promise to pay for my dinner but then ask me to contribute to paying the bill, I do not normally need to refer to general moral principles or the law to demand that you pay the entire amount (unless you have an excuse). I can simply say that I want you to pay the bill because you promised it to me. In the worst case, we simply stop cooperating and if the other person has an interest in such cooperation, this person will have the incentive to keep their promise. In smaller communities (e.g., certain hunter-gatherer communities), this individual authority is arguably sometimes sufficient to establish a fair egalitarian society (Thomas, 2006).

Finally, interpersonal entitlements are always "directed". This means that they are always equivalent to the directed duties of individuals who owe their fulfillment to us. These individuals would wrong us if they failed to do so, in which case we have an entitlement to an excuse or apology (Darwall, 2013; Wallace, 2019; May, 2015). This is not the case at least for many legal and general moral obligations (those that are not interpersonal). I might have a legal obligation not to work in China without a special permit, but I do not owe it to anyone in particular not to do it. My legal rights to own a home or to marry do not depend on any particular other individual or group of individuals in the sense that these individuals owe it to me to provide me with a home. We can easily think of having such rights without having to think of any individual group or individual as bearing a duty toward us. Similarly, my legal obligation to pay taxes at the end of the year does not depend on anyone in particular just as I have a general moral obligation, period, not to kill or steal, which does not depend on any individual in particular.

The classic concept to illustrate the notion of directedness is, again, the concept of promises. If I promise you to help your mother with her luggage and you accept (implicitly or explicitly), you have an entitlement against me that I help your mother. Importantly, there is a conceptual relationship between such entitlements and duties toward specific individuals. We cannot think of an entitlement *against* another person without ascribing a duty to them (Hohfeld, 1917). For example, we cannot think of your entitlement against me to help your mother without ascribing a duty to me (toward you) to help your mother, just as we cannot think of my duty toward you without thinking that you have an entitlement against me.

4. Networks of conceptually equivalent entitlements and other additional remarks

Any single normative relation between two or more people entails a network of many other conceptually equivalent entitlements and duties. This fact helps us to sharpen our understanding of the phenomenon of normative directedness. My duty to help your mother with her luggage entitles you to an excuse or apology if I

fail to fulfill my duty to you. It entitles you to sanction and criticize me if I breach my duty toward you and a breach entitles you to resent me. No stranger and not even your mother is entitled to me helping her based on my promise to you. Even your mother is not entitled to demand the action from me as I only gave the promise to you. It is none of her business in a way, even if she may be the main beneficiary of my promised action. It would even be inappropriate of your mother to resent me for not helping her as I do not owe it to her, e.g., because I never promised it to *her*. I only promised it to her son (cf. May, 2015). If I do more than I was supposed to (supererogation), this might make praise from *you* appropriate, but it might be odd to receive such praise from outsiders.

Even though the concept of directed or relational entitlements and commitments may be unfamiliar to some in the debate on robot rights, it is a concept that is very common in social ontology and legal philosophy (Wenar, 2021; Bratman, 2013). For example, Clodic et al. (2019) are among those who link this debate to robot ethics (see also Löhr, 2022b). That interpersonal entitlements in joint action, and interpersonal relationships more generally, are directed is rather uncontroversial (Searle, 2010; pp. 148, 177; Gilbert, 2013, p. 306; Roth, 2017). In fact, the concept of directedness (May, 2015; Darwall, 2013; Wenar, 2013) originally partly originated in legal philosophy (c.f., Hohfeld, 1917).

None of this means that all legal rights are directed. In fact, it is sometimes thought that the difference between directedness and non-directedness marks one difference between private law and public law (Wenar, 2021). The former are certain codified interpersonal rights while the latter are rights that regulate our relation to the community as a whole (as opposed to the relationship between individual members of this community). Still, as argued above, there are many more interpersonal entitlements in ordinary interactions than there are legally recognized rights. Again, if I tell you that the Sun is shining and you find out that this is a lie, it would be inappropriate for you to sue me. Still, you are entitled to sanction me in other ways, e.g., by not believing me in the future or by gossiping about me. Interpersonal entitlements vastly outnumber legal rights.

Note that my suggestion of linking the robot rights literature to the literature on relational entitlements in joint action is different from recent “relational theories” of robot rights (Coeckelbergh, 2010; Gunkel, 2018) and also from behaviorist theories (Danaher, 2020). These theories argue that we should grant moral, civil, or human rights to robots as soon as we relate to them in a certain way, whether or not they possess certain internal properties that are often thought to ground or justify certain rights. So, we are justified in ascribing rights to robots as soon as we intuitively do so.

While my account is, to a degree, compatible with these “relational accounts” – as, like them, it does not require any proof of sentience or consciousness – I build here on a different kind of relational ethics that is historically grounded not in the phenomenological tradition but in the work of German-speaking philosophers like J.G. Fichte (2005/1798), Martin Buber (1970/1923), or more recently by Stephen Darwall (2013) or Jay Wallace (2019). I take it that even skeptics in the literature on robot rights (e.g., Müller, 2021) – those who vehemently disagree with robot rights enthusiasts like Gunkel, Coeckelbergh, and Danaher – could be convinced that such entitlements will probably be possessed by highly cooperative non-sentient robots even in the next 10 or 20 years or so. Moreover, we do require certain internal properties for the notion of an interpersonal entitlement to be applicable, as we will see.

In the next two sections, I want to convince you that robots will have interpersonal entitlements against us as soon as we truly cooperate with them. This also means that we will have directed duties toward them that correspond to interpersonal entitlements. The claim is also that these interpersonal entitlements are of the *same kind* as the ones we humans have against one another in joint actions. To present my argument, I rely on the notion of “standing to make demands” (cf., Gilbert, 2018). To have an entitlement against another person is to have a certain kind of standing to make demands. A standing to make demands entails that the person of

whom something is being demanded will have a sufficient reason to act accordingly (all else being equal). The key question is whether robots could have “standing” to make legitimate demands (which would entail interpersonal entitlements). I argue that robots will have this standing if they are able to refuse to cooperate with us (an internal property) and if they are able to communicate their conditions for cooperation.

5. Where a standing to make demands come from

The question of whether robots (including chatbots) can have interpersonal directed *entitlements* against certain other humans (or perhaps even other robots) boils down to the question of what facts give human agents the standing to make demands, criticize, make requests, or waive their entitlement to an excuse. If what gives the human holder of entitlements standing to make demands and to criticize others in a joint action are sophisticated mental states, then maybe robots might never possess such standing (cf., Müller & Löhr, under contract). At least, I do not want to assume such sophisticated mental states (or any mental states) for the purpose of this paper. I do not rule them out either, but this is not what this paper is about. I argue that what gives human agent A such standing in a joint action is instead the intention of human agent B to cooperate with A. This one intention is all we need to ground the potential for *standing* or *entitlement* of A to demand that A's conditions for such cooperation are recognized and met. It gives A a kind of “bargaining power”.

Note that I do not claim that my proposal here is the only way entitlements in interpersonal relationships can be grounded. I merely argue that it is one way standing or entitlement can be grounded and that this is a ground that can be ascribed to robots. It is a story of how interpersonal entitlements might arise in interactions that make sense in the case of our interactions with robots. I do not take this to be the end of the discussion and rather consider this a kind of impetus to the debate where it is sometimes denied that the notion of robot rights even makes sense or is conceivable.

The account I put forth has two elements – a conceptual analysis of the notions of valid demands and requests that are partly inspired by the literature on relational ethics (especially Darwall, 2013). This is what I will do in this section. In the next section, I introduce the second element. I show how these concepts apply to phenomena that naturally arise from our cooperative interactions. Finally, in the last section, I show how these concepts can be applied to robots.

The first element of this account is a conceptual analysis of the concept of a valid demand in terms of the notion of reason. I assume that a valid demand of the form “I demand of you to ϕ ” conceptually entails that making such a demand gives the person we demand something of a *sufficient reason* (all else being equal) to act according to the demand (cf., Darwall, 2013). Importantly, this person has a sufficient reason to act even if this goes against their self-interest (Gilbert, 2018 for a similar assumption). If this demand was not legitimate, then the person making the demand would not have the standing to make it (this is just what it means to make an illegitimate demand). Thus, the person who is the addressee of the demand has no (sufficient) reason to act on it (all else being equal).

This little conceptual analysis is all I need to make my case for robot entitlements. Again, I simply take this to be a claim about our concept of demands. A valid demand is such that the person to whom the demand is addressed must act accordingly (all else being equal). Compare the notion of a legitimate demand to the notion of a legitimate request. A legitimate request does *not* give the other person sufficient reason to act (contrary to the order or the demand). Instead, it gives the other person the entitlement to accept or reject the request. It is up to us whether we wish to act on a request or not. If it isn't up to us, we would have to do with an order or a legitimate demand – a different speech act.

To illustrate this conceptual analysis further, compare both a demand and a request with the concept of an order. A legitimate order also gives the other person sufficient reason to act according to the order (if there is

sufficient standing for it). The difference between an order and a demand is that there was no entitlement to the action before the order was made. The person who has the authority to make orders is able to create duties to act (see Darwall, 2013 for this analysis of the concept of order). An order is different from a request in the sense that the person who is ordered to do something does not have the standing to reject the order (if it was a legitimate order) and no justification for the ordered action is needed.

For example, if I demand of you to keep your promise, and this demand is legitimate (I had the standing to make it), then this means that you have a *sufficient reason* to keep your promise even if you don't feel like it anymore (all else being equal). You may ask me to waive your promissory right against me or to excuse your lack of willingness to keep the promise, but all of this entails is that my demand was legitimate. And this means that I am entitled to the demanded action.

Similarly, if we are taking a walk together and you walk too fast for me, I can *ask* you to slow down. In this case, you can decide whether you want to follow my request or not. However, I can in some cases also demand of you to slow down if I was entitled to you walking at the same speed. If this demand is legitimate, then you have sufficient reason (all else being equal) to slow down, whether you want to or not. If my demand gives you a sufficient reason to slow down. This, again, shows that I was entitled (against you) that you slow down. If my demand was illegitimate, you may point this out to me, and I am normally required to apologize for making an illegitimate demand. I can then still ask you to slow down (again, all else being equal).

Note that 'sufficient reason' here simply means that (all else being equal) no other reason or justification is needed to make the case that you owe it to me to, say, slow down. If I demand of you to slow down and this was a valid demand, which entails that I am already entitled to you slowing down, then you do not need any more justification for slowing down. Similarly, if my order was legitimate and, thereby, creates a duty (because I have the relevant authority), then this duty gives you a sufficient reason (all else being equal, i.e., unless there are conflicting reasons) to act according to your duty.

For example, imagine you are taking a walk and you walk too slowly for your partner to keep up. When she demands of you to slow down and you recognize the legitimacy of this demand (assuming it was legitimate), then you do not need any more reason or justification to know what you ought to do or what you owe your partner. Her demand is enough reason for you to slow down – even if you do not want to. You would arguably wrong her if you demanded an additional justification.

6. How a standing to make demands can arise from our interactions

The second element of this account emerges from reflections on our joint activities with others (Löhr, 2022 b). Whenever you are engaged in an action with another person (whether it's buying coffee from your favorite barista or taking a walk with a friend), you have sufficient reason to think that this person intends to cooperate with you in some way or another (i.e., no additional justification or reason is needed). All this means is that if this person did not have any such intention, she would not interact with you and would simply leave. Even uncooperative agents intend to interact with you at least for the time being and even if they intend to eventually mislead you (be uncooperative). For you to be able to lie to me (assert something you believe is false to deceive me) requires that you are at least to some degree interacting with me, even if it is just to tell me the lie. And this entails that you intend to interact with me, at least for now. In other words, any joint action entails that the agents that are cooperating intend to cooperate in some way or other at least for the time being.

Note that forming an intention gives the agent sufficient reason to act on this intention (an intention is decisive, all else being equal, cf. Bratman, 1999; Gilbert, 2013). This insight gives the cooperating agents sufficient reason to believe that everyone who is interacting in the group has sufficient reason to act cooperatively to some degree at least (because this will fulfill their individual intentions to cooperate). The important idea now is that

in a joint action, the fulfillment of my intention to cooperate does not just depend on me *but also on you*. For my intention to be fulfilled, you will have to be motivated to cooperate with me. This gives the other person a kind of *standing* to make *their* conditions for cooperating explicit. So, if I have sufficient reason to think that another person intends to cooperate with me in (some way or another), this automatically gives me a certain kind of *standing* to make my conditions for cooperating with this person explicit to them. I have reason to expect that this person has either sufficient reason to consider and accept my conditions for cooperation or to change their mind about cooperating with me.

The important step in the argument now goes back to the analysis of the concepts of valid demands and requests. “Standing to make a demand or request” (meaning that my demand or request is legitimate) entailed that the other person has sufficient reason to act accordingly. So, if I intend to cooperate with another person, this gives me sufficient reason to act cooperatively. If acting cooperatively means acting in accordance with the other person’s conditions for cooperation (to some degree at least), then this means that I have sufficient reason to act according to the conditions of cooperation of the other person, when these conditions are voiced. This at least looks a lot like a valid demand grounded in the intention of the person who intends to engage in a joint action. Moreover, it follows from the analysis of the concept of a valid demand. Put more simply, I have a certain kind of bargaining power in a joint action because it also depends on *me* whether the other person’s intention to cooperate with me will be fulfilled.

My argument is best explained using an example from the literature in social ontology (e.g., Gilbert, 2013). Imagine that Sarah and Arnold are taking a walk together (as it is a “nice night for a walk”). Both Sarah and Arnold have reason to believe that the other intends to cooperate with the other – in this case, to take a walk together. This is simply entailed by their interaction. If Arnold did not have the intention to take a walk with Sarah right now when they are in fact taking a walk together, he would not take a walk with her. From this assumption, it follows that Sarah has standing to make her conditions for such cooperative activities explicit to Arnold. Some of these conditions are such that Arnold will have sufficient reason to act accordingly, given that he intends to sustain the cooperative interaction with Sarah. The conditions that give him sufficient reason to fulfill are such that we can consider them grounded in Sarah’s entitlement to demand them of Arnold. This “sufficient reason” simply derives from his intention to cooperate with Sarah, assuming it is a condition that has to be met for him to fulfill his intention to cooperate with Sarah.

Imagine now that Sarah *demands* of Arnold to slow down if he walks too fast. This is her condition for further cooperating with Arnold. If this demand is legitimate – if Sarah had the proper standing to make it – Arnold now has sufficient reason to slow down even if he himself wants to walk faster. Again, for this reason, to be a *sufficient* reason, it has to be the case that asking for a justification would be inappropriate (he would not be entitled to it), such that Sarah will respond “Because I do not want to walk fast, this should be reason enough.” Of course, it might be that the demand was illegitimate and that Sarah had no such standing – maybe she only had an entitlement to make a request. In this case, Arnold might object: “Sarah, you know, you can also speed up! Why don’t you speed up? Why do I have to slow down?” They might then negotiate the adequate speed that they are both comfortable with. A standing for making a request also conceptually entails an entitlement. If this request is valid, this entails that the other person has a duty toward the requesting person to respond to the request. Thus, we need not always talk about demands, we can also talk of more mundane entitlements, like entitlements to make requests or notify when we want to leave. The idea is that our actions are constantly constrained by all kinds of implicit interpersonal entitlements and duties.

Again, let me emphasize that the standing to make demands does not derive from any laws or even from moral principles. It simply derives from the person’s intention to cooperate with another person. If Arnold wants to cooperate with Sarah, and Sarah’s condition for doing so means that Arnold has to slow down, he now has a choice. Either he changes his mind and no longer has the intention to cooperate with Sarah, or he complies. A

third option is again to argue that the demand was not legitimate but that he also has the standing to negotiate different terms of conditions for cooperation. Now, if Sarah also has the intention to cooperate with Arnold, Arnold most likely has the necessary standing to make his conditions for cooperation explicit. If the conditions overlap sufficiently, they most likely are both motivated to continue the joint activity. If his conditions for cooperation are not fulfilled, Arnold can refuse to keep cooperating with Sarah and the joint action disintegrates and his personal intention will not be fulfilled. If you act in ways that I find so unacceptable that I lose motivation to cooperate with you, then I will simply walk away.

I take it that most readers have been in a situation with a friend, colleague, or partner where we establish certain basic conditions for future cooperation that the other person has only reason to accept because they want to keep cooperating with us in the future. Healthy romantic relationships seem to be made exactly of this kind of glue of reciprocity. I want my partner to respect my interests not for moral or legal reasons but simply because they want our relationship to continue, i.e., for both of us to be sufficiently motivated to keep going. Again, note that I do not argue that this account captures everything we want a moral theory of rights in the moral or legal domain to capture. But this is not my ambition here. I do not aim at a general account of demand rights. I want to develop one reasonable foundation of a notion of entitlement that may apply to robots who themselves might not have sophisticated mental states. The aim is to develop a plausible and reasonable account of standing to make demands (which entails entitlements) that can be attributed to robots.

So, in summary, valid demands, requests, and orders are actions that give the other person sufficient reason to act accordingly. To make a valid demand or request entails standing or entitlement. In the case of a valid demand, the reason for action had already been there and is simply made explicit. In the case of a request, only a sufficient reason to accept or decline the request with a justification is in order. In interactions, we have sufficient reason to believe that all parties intend to cooperate to some degree at least (otherwise the interaction would disintegrate). An intention to cooperate gives the agents sufficient reason to act cooperatively. To act cooperatively means to act such that all agent's conditions for cooperation are met. Since we do not always know what these conditions are, we often make them explicit. Sometimes, communicating our conditions gives the other agent(s) sufficient reason to act accordingly (no further justification is needed). Other times, they may have sufficient reason to negotiate (the conditions are not yet accepted). In the former case, we can speak of a valid demand. In the latter case, we can speak of a valid request. Both entail entitlements.

7. Why social robots can have interpersonal entitlements in a joint action

Sophisticated robots (including chatbots) who will truly cooperate with us will have the same standing to make demands as human agents if they come equipped with an important feature: the ability to refuse cooperation if certain conditions are not met. The conditions under which they will refuse to cooperate with another person can be programmed into it. We do not need to assume that these conditions must arise somehow without explicit training or without building such a capacity into the robot. Consequently, we do not need to posit any real autonomous decision-making or any real intentions, consciousness, or anything that normally justifies moral or legal rights. Again, I am not talking here about general legal or moral rights (rights that we do not have against specific individuals) and instead focus on a kind of right or entitlement that naturally arises from interpersonal interactions. The cognitive requirements for robot standing in joint action are quite high but they are also quite low in the sense that it requires nothing that cannot be built into a social robot or chatbot arguably even in the next ten or twenty years. Robots don't require any "real interests" or consciousness for them to have standing in the sense defined here. All they need is somewhat sophisticated behavior. For all we know, they could be *philosophical zombies* (cf., Chalmers, 1996, for this concept).

How exactly does this account of robot rights work? The idea is simple. Imagine that Arnold is a robot and Sarah is a human being. If human Sarah intends to interact with robot Arnold and Arnold has conditions for such an intention to be fulfilled, then Sarah has sufficient reason to act such that Arnold's conditions are fulfilled. This

follows from the idea that an intention gives us sufficient reason to act according to this intention (all else being equal, e.g., Bratman, 1999). In other words, if Sarah has an intention, this is normally reason enough for her to act accordingly (all else being equal), i.e., she does not need any further justification (cf., again Gilbert, 2013). Even if the conditions of the robot are not compatible with her immediate interests, if they need to be met for her general intention to cooperate to be fulfilled, she still has sufficient reason to fulfill them (otherwise she will simply change her mind and not intend to cooperate with the robot). So even if Sarah might not like to walk too fast, if this is a condition for taking a walk at all, Sarah might still have reason to comply even if it means walking a bit faster. If Arnold did not have the standing to make demands, we can imagine that at least he had the standing to make a request and Sarah had a sufficient reason to respond to this request, assuming that this is one of the robot's conditions for cooperation. This entitlement (to demand or request) simply derives from Sarah's intention to cooperate with Arnold, which gives Arnold standing to make his conditions for Sarah's intention to be fulfilled explicit.

Let me put this idea more simply: If you intend to cooperate with a robot and you know that the robot shuts down under certain conditions, you will try to act in ways that the robot is programmed to consider acceptable for the continuation of the joint activity. In fact, you even have sufficient reason to do so simply because you intend to keep cooperating with the robot and not necessarily because you have a general moral or legal duty to do so or because you care about the interests of the robot. Just as human beings have their boundaries, so will many robots come equipped with conditions for cooperation. Just as we often do not know where the other's boundaries lie, it is also possible that we do not know this of every robot we interact with in the future. In this case, we have to ask or react appropriately to criticism. Arguably, this lack of access to others' conditions for cooperation is one reason for why we make our demands explicit to them even if we are already entitled to the demanded action. We know that if people or robots lose interest in cooperating with us (i.e. if their conditions for cooperation are not met), they will likely withdraw from the joint action and the joint action disintegrates – our intention to cooperate will not be satisfied.

To further understand the possibility of a normative relationship between a robot and a human, we should step away from focusing on short interactions and turn to interactions with the same robot over a longer period of time. Think of the robot Arnold as a service robot that lives with a family and goes about his daily business of doing the dishes, cooking, and protecting the property of the family, etc. John, Sarah's teenage son, intentionally bumps into Arnold. Arnold's algorithm is designed to explain to John that this is not okay. Moreover, if John wants Arnold to prepare him a meal and help him with his homework, i.e., cooperate with him, he should not act in this way for the simple reason that it will cause his cooperation to end. Arnold is programmed to retain information about every significant event and if John fails to change his behavior, Arnold will stop cooperating with him, i.e., ignore him. For example, if John tells Arnold to prepare fresh orange juice for him, Arnold will ignore this request. Arnold will not justify the refusal to cooperate. He doesn't have to. It is his condition for cooperation. And the existence of conditions for cooperation is the most fundamental basis for the possibility of making a legitimate demand. If John wants to keep cooperating with Arnold, he has sufficient reason to stop pushing him (which is how I defined the notion of standing to make demands above) and to even apologize (signal compliance with Arnold's conditions). In other words, Arnold's demands were valid. He had standing – he was entitled to make them!

What I would like to further insist on is that the entitlements I wish to ascribe to Arnold and other robots of his kind are *directed*. The entitlement in question is directed at the other collaborators. They cannot be reduced to the human owner or the creator of the robot. We can see this by showing that the conditions for directedness specified above (*specificity* and *equivalence*) are met. First, we can observe that the specificity requirement is met. When John pushes Arnold, it is not that he wrongs the company that builds Arnold or even John's parents. Arnold did not get damaged by John and no property law was violated. The wrongness exists merely in the relation between him and Arnold. When John wants to repair his relationship with Arnold (e.g., because he

wants Arnold to make a sandwich), he apologizes to him and not the company that rents Arnold out. An apology to the company or his parents would not have any function, at least not the same function of repairing his relationship with Arnold. The apology must therefore be addressed to Arnold (for more on this idea of functions of apologies see Löhr, 2022 b). Criticism, too, must be addressed to Arnold as opposed to the parents (the owners) or the company. John might have the standing to correct Arnold's behavior by providing negative feedback ("You should not add mayonnaise, Arnold"). If John complained about Arnold's behavior to the producer of the robot, the producer would be entitled to point out that Arnold's algorithm still has to learn certain things. They would be responsible for a defective robot, but Arnold is not defective. He works as he should.

Second, we can observe a certain kind of *conceptual equivalence* between the kind of actions John is entitled/obligated to and the kind of actions robot Arnold is entitled/obligated to, which cannot be observed in the relationship between the robot and the company that owns or built the robot. If Arnold and John were to play a game, for example, it is not unreasonable to think that John is entitled (against Arnold) that Arnold play by the rules of chess. This also means that Arnold is obligated (toward John) to play by the rules. However, John does not have this entitlement against the company that rented Arnold out or the parents who own Arnold. When John complains that Arnold fails to play by the rules, both the parents or the company might respond that there is nothing they can do and advise John to keep playing so that Arnold learns the rules of the game and also what John expects of Arnold. It seems that John has an entitlement against Arnold that cannot be thought of without thinking of Arnold owing it to John to play by the rules. John's entitlement is not conceptually equivalent to any duty possessed by the parents or the company who do not owe John that Arnold plays by the rule (unless, e.g., they promised John that their product would have this feature, but this would be an entirely different normative relationship).

So, again, Arnold's duty toward John to play by the rules is conceptually equivalent to John's entitlement against Arnold to play by the rules. It is not conceptually equivalent to any entitlement the company might have (e.g., not to damage Arnold or to pay the monthly rental fee). The company is not entitled against John that John is playing by the rules of chess. Arnold, however, is when he says that unless John plays by the rules, he will simply refuse to play with him in the future. Again, the set of entitlements the company has, and the set of entitlements Arnold has is completely different. It is simply not plausible to say that it is not John who has an entitlement against Arnold that Arnold plays by the rules when playing chess with him, but that it is a right that the company that owns Arnold has against John. John does not owe it to the company and would not wrong them if he were to break the rules when playing chess with Arnold. So, in summary, it seems that Arnold not only has standing to make demands to whoever wants to cooperate with him but that this standing is directed in nature. It is not something that is indirectly grounded in the right of the owner of the robot or anyone else. It is grounded in the cooperative relationship between those who have intentions to cooperate with the robot.

8. Conclusion

I showed that robots (including chatbots) without consciousness can have the standing to make demands, which entails that robots can have entitlements, i.e., a kind of rights. Moreover, such entitlements are not derived from the rights of humans. They are grounded in the interaction between the human and the machine – this means they are directed and interpersonal. This is not a relational account in Gunkel's (2018) sense, for example, as I require an important internal property of robots to ground rights – they must have the capacity to refuse to cooperate. It is a relational account in the sense of relational ethics derived from the work of German idealists like Fichte or Buber. In a nutshell, if a robot can refuse to cooperate, then it acquires a kind of bargaining power. If there is a human being who intends to cooperate with the robot, then the robot can make its conditions for cooperation explicit. I showed that this gives the human agent sufficient reason to act according to the conditions. I have argued that this is exactly what we would expect from a legitimate demand. Once robots

have acquired this bargaining power, they might acquire sufficient standing to demand legal protection, e.g., not to be abused or turned off. But exploring this last conjecture will be left for future work.

References

- Basil, 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* (pp. 289–306). Oxford University Press.
- Bratman, M. E. (1999). *Faces of intention: Selected essays on intention and agency*. Cambridge University Press.
- 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) (pp. 63-74). Arnold Benjamins Publishing. <https://doi.org/10.1075/nlp.8.11bry>
- Buber, M. (1970). *I and Thou*. Simon and Schuster.
- Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford Paperbacks.
- Coeckelbergh, M. (2010). Robot Rights? Towards a Social-Relational Justification of Moral Consideration. *Ethics and Information Technology*, 12(3), 209–221. <https://doi.org/10.1007/s10676-010-9235-5>
- Danaher, J. (2020). Welcoming robots into the moral circle: a defence of ethical behaviourism. *Science and Engineering Ethics*, 26(4), 2023–2049. <https://doi.org/10.1007/s11948-019-00119-x>
- Darling K. (2012) Extending legal rights to social robots. *We Robot Conference*. University of Miami, April. Retrieved 8 April 2014 from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2044797
- Darwall, S. (2013). *Morality, authority, and law: Essays in second-personal ethics I*. OUP Oxford.
- Fichte, J. G. (2005). *Fichte: The System of Ethics*. Cambridge University Press.
- Geller, J. C. (2020). *Rights for Robots: Artificial Intelligence, Animal and Environmental Law (Edition 1)*. Routledge.
- Gerdes, A. (2016). The Issue of Moral Consideration in Robot Ethics. *ACM SIGCAS Computers and Society*, 45(3), 274–279. <https://doi.org/10.1145/2874239.2874278>
- 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.
- Gomez-Lavin, J., & Rachar, M. (2019). Normativity in joint action. *Mind & Language*, 34(1), 97–120. <https://doi.org/10.1111/mila.12195>
- 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>
- Gunkel, D. J. (2018). *Robot rights*. Cambridge: MIT Press.
- Gunkel, D. J. (2020). 2020: The Year of Robot Rights. *The MIT Press Reader*. Retrieved from <https://thereader.mitpress.mit.edu/2020-the-year-of-robot-rights/>
- Hohfeld, W. N. (1917). Fundamental legal conceptions as applied in judicial reasoning. *The Yale Law Journal*, 26(8), 710–770. <https://doi.org/10.2307/786270>
- 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. <https://doi.org/10.2139/ssrn.3762223>
- Levy, D. (2005). *Robots unlimited: Life in a virtual age*. CRC Press. <https://doi.org/10.1201/b10697>
- Löhr, G. (2022a). Recent experimental philosophy on joint action: Do we need a new normativism about collective action? *Philosophical Quarterly*, 72(3), 754–762. <https://doi.org/10.1093/pq/pgab070>
- Löhr, G. (2022b). Robot Rights in Joint Action. In Müller, V.C. (Ed.) *Philosophy and Theory of Artificial Intelligence 2021. PTAI 2021. Studies in Applied Philosophy, Epistemology and Rational Ethics*, Vol 63. Springer, Cham, pp. 179–191. https://doi.org/10.1007/978-3-031-09153-7_15
- Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and information technology*, 6, 175–183. <https://doi.org/10.1007/s10676-004-3422-1>
- May, S. C. (2015). Directed duties. *Philosophy Compass*, 10(8), 523–532. <https://doi.org/10.1111/phc3.12239>
- Miller, L. F. (2015). Granting automata human rights: Challenge to a basis of full-rights privilege. *Human Rights Review*, 16(4), 369–391. <https://doi.org/10.1007/s12142-015-0387-x>
- Müller, V. C. (2021). Is it time for robot rights? Moral status in artificial entities. *Ethics and Information Technology*, 23, 579–587. <https://doi.org/10.1007/s10676-021-09596-w>
- Müller, V. & Löhr, G. (under contract). *Artificial Minds*. Cambridge University Press.
- Nyholm, S. (2020). *Humans and robots: Ethics, agency, and anthropomorphism*. Rowman & Littlefield Publishers.
- 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. <https://doi.org/10.4324/9781315768571>
- Searle, J. (2010). *Making the social world: The structure of human civilization*. Oxford University Press.
- Schwitzgebel, E., & Garza, M. (2015). A defense of the rights of artificial intelligences. *Midwest Studies in Philosophy*, 39, 98–119. <https://doi.org/10.1111/misp.12032>
- Schwitzgebel, E. (2023). The Full Rights Dilemma for AI Systems of Debatable Moral Personhood. *ROBONOMICS: The Journal of the Automated Economy*, 4, 32. Retrieved from <https://journal.robonomics.science/index.php/rj/article/view/32>
- 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.
- Sparrow, R. (2007). Killer robots. *Journal of Applied Philosophy*, 24(1), 62–77. <https://doi.org/10.1111/j.1468-5930.2007.00346.x>

- Strawson, P. F. (2008). *Freedom and resentment and other essays*. Routledge. <https://doi.org/10.4324/9780203882566>
- Tigard, D. W. (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>
- Tigard, D. W. (2021). There is no techno-responsibility gap. *Philosophy & Technology*, 34(3), 589-607. <https://doi.org/10.1007/s13347-020-00414-7>
- Thomas, E. M. (2006). *The old way*. New York: Farrar, Straus & Giroux.
- Turner, J. (2019). *Robot Rules: Regulating Artificial Intelligence*. Berlin: Springer. <https://doi.org/10.1007/978-3-319-96235-1>
- 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(4), 375–399. <https://doi.org/10.1163/174552412X628968>
- Wenar, L. (2021). Rights. In Zalta, E. N. (Ed.) *The Stanford Encyclopedia of Philosophy*. Retrieved from <https://plato.stanford.edu/archives/spr2021/entries/rights/>

Received: 27/12/2022

Revised: 16/09/2023

Accepted: 19/10/2023