Human-Robot Moral Relations: Human Interactants as Moral Patients of their Own Agential Moral Actions Towards Robots

Cindy Friedman[1,2][0000−0002−4901−9680]

[1] Department of Philosophy, University of Pretoria, Pretoria, South Africa
cindzfriedman@gmail.com
[2] Centre for AI Research (CAIR), South Africa

Abstract. This paper contributes to the debate in the ethics of social robots on how or whether to treat social robots morally by way of considering a novel perspective on the moral relations between human interactants and social robots. This perspective is significant as it allows us to circumnavigate debates about the (im)possibility of robot consciousness and moral patiency (debates which often slow down discussion on the ethics of HRI), thus allowing us to address actual and urgent current ethical issues in relation to human-robot interaction.The paper considers the different ways in which human interactants may be moral patients in the context of interaction with social robots: robots as conduits of human moral action towards human moral patients; humans as moral patients to the actions of robots; and human interactants as moral patients of their own agential moral actions towards social robots. This third perspective is the focal point of the paper. The argument is that due to *perceived* robot consciousness, and the possibility that the immoral treatment of social robots may morally harm human interactants, there is a unique moral relation between humans and social robots wherein human interactants are both the moral agents of their actions towards robots, as well as the *actual* moral patients of those agential moral actions towards. Robots, however, are no more than *perceived* moral patients. This discussion further adds to debates in the context of robot moral status, and the consideration of the moral treatment of robots in the context of human-robot interaction.

Keywords: Robot Ethics, Human-Robot Interaction, Moral Patiency.

1 Introduction

This paper contributes to the debate in the ethics of social robots on how or whether to treat social robots morally by way of considering a novel perspective on the moral relations between human interactants and social robots: that human interactants are the *actual* moral patients of their agential moral actions towards robots; robots are no more than *perceived* moral patients. This novel perspective is significant because it allows us to circumnavigate contentious debates surrounding the (im)possibility of robot consciousness and moral patiency, thus allowing us to address actual and urgent current ethical issues in relation to human-robot interaction (HRI).

Social robots are becoming increasingly sophisticated and versatile technologies. Their wide range of potential utilisations include carer robots for the sick or elderly (see e.g. [50] [53] [60], general companion robots (see e.g. [15] [57]), teachers for children (see e.g. [36] [51]), or (still somewhat futuristic but nonetheless morally relevant in human-robot interaction (HRI) contexts) sexual companions (see e.g. [21] [37] [48]).

Although social robots may take on a variety of forms – such as the AIBO robot who takes the shape of a dog, or the Paro robot that takes the shape of a baby seal – I will here be focusing on android social robots.[1] This is the case as a combination of a human-like appearance and human-like sociability creates the potential for human interactants to relate to these robots in seemingly realistic human-like ways.

Given the possibility for human interactants to relate to these social robots in human-like ways[2], researchers have investigated not only the nature of these relations and how they may morally impact us – Turkle [57], for example, puts forward that some relations with robot companions may fundamentally change what it means to be human, and Nyholm & Frank [48] speculate that certain relations with robots may hinder us from forming bonds with other people – but also whether we have a *moral* relation to these robots that would require us to relate to them in a *particular way.* By this, I mean - should we treat them morally well? For example, someone such as Bryson [12] argues vehemently against the need for moral treatment of robots, whereas some, such as Levy [38] or Danaher [18], argue in various ways that we should consider the moral treatment of robots.

This paper will consider the issue of the moral treatment of social robots from an *anthropocentric persective* (as opposed to a 'robot perspective') by considering arguments that treating a robot immorally causes moral harm to its human interactant. Given this possibility, I suggest that in this context, social robots and human interactants have a unique moral relation: human interactants are both the moral agents of their actions towards robots, as well as the *actual* moral patients of those agential moral actions towards. Robots, in this case, are no more than *perceived* moral patients.

Literature on robot ethics is less focused on patiency as it is agency (with regard to both human interactants and robots in the HRI context) (see e.g. [38] [27]), and where there is a focus on patiency as far as robots are concerned, it most often discusses the notion of the moral treatment of robots from the perspective of the current (im)possibility for robots to be *actually* conscious and,

[1] Unless otherwise specified, any use of the term 'social robot' will specifically refer to android social robots.

[2] It must be noted that social robots cannot genuinely reciprocate human sentiments; they cannot care for a human interactant the way in which a human interactant may care for them (e.g. [16]). Any emotions displayed by robots are functional in nature, thus, at least currently (or even in the near future), human interactants cannot have genuinely reciprocal or mutual bonds with robots (e.g. [55]). Thus, any relation or bond formed with a social robot is unidirectional in nature.

thus, the (im)possibility for them to be *actual* moral patients (see section 4). However, in putting forward that it is human interactants who are the moral patients of their own agential moral actions towards robots, we may circumnavigate the somewhat intractable debate of *actual* robot consciousness which arises in relation to the (im)possibility for robots to be moral patients in the context of questioning whether they warrant moral treatment. This is not to say that concerns surrounding artificial robot consciousness are unimportant, but rather to say that we should not become so detained by the concern as to whether robots can be conscious or not (and thus moral patients or not) that we are misdirected from addressing actual and urgent current ethical issues in relation to human-robot interaction. My argument that it is human interactants who are the actual moral patients of their agential moral actions toward social robots thus allows us to seriously consider these actual and urgent current ethical issues.

I will first discuss two instances wherein human interactants are moral patients in relation to the robots with which they interact: firstly, robots as conduits of human moral actions towards other human moral patients; secondly, humans as moral patients to the moral actions of robots. I will then introduce a third perspective wherein a human interactant is, at the same time, both a moral agent and a moral patient: human interactants as moral patients of their own agential moral actions towards robots. I will firstly distinguish between the *actuality* of robot consciousness and the *perception* of robot consciousness since this is important for our understanding of robots as *perceived* moral patients, and also for our understanding of why, in the context of this paper, the *actuality* of robot consciousness is a non-issue. I will then put forward that treating social robots immorally may cause moral harm to human interactants and I do so using three sub-arguments: social robots are more than mere objects; the act of treating a social robot immorally is abhorrent in itself; and, due to these arguments, treating a social robot immorally may negatively impact upon the moral fibre of interactants. Finally, due to the perception of robot consciousness, and, thus, the perception of robot moral patiency, as well as concern that treating social robots immorally may cause moral harm to human interactants, I argue a human interactant is, at the same time, both the agent and patient of their moral actions towards robots: human interactants are the *actual* moral patients of their agential moral actions towards robots, whereas robots as *perceived* moral patients.

Let us now consider two ways in which human interactants may be moral patients in the context of their interaction with robots so as to contextualise the argument this paper makes, and make clear how and why my contribution is a particularly novel one.

2 Robots as conduits of human moral action towards human moral patients

Although this category of human moral patiency is related to computer ethics, it can also be applied to robot ethics. Regarding this first distinction, computer ethics, for example, "endeavors to stipulate the appropriate use and/or misuse of technology by human agents for the sake of respecting and protecting the rights of other human patients" [27]. We may consider the first commandment in the *Ten Commandments of Computer Ethics* [5]: "Thou shalt not use a computer to harm another person." And, more recently, the Institute of Electrical and Electronics Engineers (IEEE) initiatives on AI ethics and automous systems (https://ethicsinaction.ieee.org/) [1]. From the perspective of computer ethics, computers are ultimately deployed by humans, used for a human purpose and, as such, have an effect on humans. An example could be using computer technology through social media to spread fake news or deface somebody's character.

Or, as far as robotics and robot ethics ('roboethics') is concerned wherein we grapple with the ethical issues of the use of robots (see e.g. [40] and [47], the possibility exists of directly commanding a robot to injure another human being. In such instances, a human agent would not be directly interacting with another human patient, they would be treating a human patient immorally through the use of technology – such as a computer or robot; technology would be the conduit of immoral action on behalf of the human agent, directed at another human patient. Although the robot is conducting the immoral action against the human moral patient, the difference (as compared to the second perspective discussed below) is that there is direct human intervention whereby the moral decision is ultimately made by a human, and the human agent uses technology to then inflict the moral harm that is the result of the decision they have made. For example, in terms of autonomous weapons systems (AWSs), there is a distinction between AWSs which "operate entirely independently of human controllers, and teleoperated unmanned weapons systems, which are still under remote human control" [20]. Teleoperated weapons systems would be a case of a human agent ultimately making a moral decision as to whether to harm a human moral patient or not, but using an AWS to carry out the decision. AWSs that operate entirely independently of human controllers would fall under the second category (see next section) relating to machine ethics – humans being moral patients to moral decisions made by technology or, particularly in this instance, robots.

3 Humans as moral patients to the actions of robots

As far as the second perspective is concerned, machine ethics (ME), for example, "seeks to enlarge the scope of moral agents by considering the ethical status and actions of machines" [27]. It (ME) "reasserts the privilege of the human and considers the machine only insofar as we seek to protect the integrity and the interests of the human being" [27]. It considers the possibility of machines to be guided by ethical principles in the decisions that it makes about possible courses of action [2]. As such, the machines in question are machines that make decisions and act autonomously (without human intervention) by way of "[combining] environmental feedback with the system's own analysis regarding its current situation" [29]. Given this understanding of autonomous decision making systems (ADM systems) that have the potential to be moral agents, we can then consider the possibility that humans can be moral patients to the moral decisions and actions of AI. Specifically, in our context, this potentiality means that robots could harm humans.

The topic of the possibility for machines to be considered moral agents is a broadly contested and complicated one, full discussion of which would go beyond the confines of this paper. However, it is worth noting some arguments that have been made concerning the topic. Generally speaking, the topic is one which questions whether machines can be moral agents – is morality programmable? – and what conditions they would have to fulfill in order to be considered moral agents, as well as the impact that these agents would have on us.

Well-known researchers weighing in on the issue include Asaro [4], Bostrom and Yudkowsky [7], Brundage [11], Deng [24], Lumbreras [41], McDermott [43], Moor [46], Sullins [54], Torrance [56], Wallach and Allen [61], Wang & Siau [62], and many others. Different sets of conditions for moral agency are suggested: A combination of free will, consciousness, and moral responsibility [61]); a combination of the abilities to be interactive, autonomous, and adaptable [25], and a combination of autonomy, responsibility and intentionality [54]. Do we need to ensure artificial moral agents (AMAs) are both ethically productive and ethically receptive [56], or is the ability for rational deliberation all that is needed [39]?

Although it is debatable whether robots can or cannot truly be moral agents given how philosophically loaded the topic is, it remains that, regardless of this uncertainty, humans can still be moral patients of the actions of autonomous machines that act without direct human intervention. For instance, and going back to the example mentioned above of AWSs, although we could debate endlessly about whether an AWS that acts without human intervention is a moral agent, the fact remains that it can still ultimately make the moral decision to kill a civilian or not, and this civilian would be the moral patient of this moral decision – whether they lived or died.

This is not to say that were the AWS to kill a civilian, it would hold full moral responsibility for the civilian's death – this is another complex issue entirely[3] – nor is it to say that the AWS is, in and of itself, a moral agent. Rather, it is to say that moral responsibility and agency aside, the civilian would have been killed due to a decision ultimately made by the AWS (although the groundwork for the decision would be based on its programming). At that moment, there is no direct human intervention wherein a human is making the decision to kill the civilian or not.

Thus, as stated above, there is the potential for human beings to be harmed by this technology.

[3] The topic of moral responsibility is also a contentious one and there remains what can be termed a *responsibility gap* when it comes to who should be held responsible for the actions of autonomous systems (see e.g. [42]).

4 Human interactants as moral patients of their own agential moral actions towards robots

I will now argue that there is a third perspective we may consider in relation to human interactants being moral patients in the context of their interaction with social robots: *human interactants as moral patients of their own agential moral actions towards social robots.* Before I can put this moral relation forward, we first need to understand the difference between the *actuality* of robot consciousness and the *perception* of robot consciousness, since this distinction is important in relation to the understanding of human interactants being the *actual* moral patients of the agential moral actions towards robots, and robots being the *perceived* moral patients of these actions. I will then briefly discuss arguments made that treating social robots immorally may morally harm human interactants. Given the *perception* of robot consciousness, and the potential that treating social robots immorally may morally harm human interactants, I investigate the unique moral relation that then arises between human interactants and social robots: that a human interactant is, at the same time, both the moral agent, as well as the *actual* moral patient, of their moral actions towards social robots – specifically in the context of immoral treatment. Social robots, however, are the *perceived* moral patients of such moral actions.

4.1 The actuality of robots consciousness vs. the perception of robot consciousness

The very topic of consciousness – what it is, and what it means to be conscious – is a hugely contested one. We still seem to be far away from having a definitive answer as to what consciousness is in the human sense, let alone what it would mean for an AI to be conscious, and whether this would ever be a possibility. How can we even begin to formulate a definitive answer in the context of artificial consciousness, when we seem no closer to understanding our own consciousness? Although I here remain agnostic to the possibility of conscious AI, and hold that we need not concern ourselves with it too much in the context of this paper, given the *perception* of robot consciousness (discussed below), it is worwhile to consider some arguments in the context of the *actuality* of robot consciousness. Doing so demonstrates the intractibility of the issue of consciousness in AI, and why I hold that it is more beneficial in the context of my arguments to circumnavigate the debate entirely

Property dualists, such as David Chalmers, make the distinction between the easy and the hard problems of consciousness. According to Chalmers [13], the easy problems of consciousness pertain to explaining the following phenomena: "the ability to discriminate, categorize, and react to environmental stimuli; the reportability of mental states; the ability of a system to access its own internal states; the focus of attention; the deliberate control of behaviour; the difference between wakefulness and sleep". If we were to artificially replicate the human brain, we would merely be creating an AI that acts *as if* it is conscious and arguably dealing at best with the easy problems. However, it is far from clear in philosophical circles that consciousness can be determined behaviouristically (see e.g. [14] [35]. As such, creating an AI that behaves the way in which a conscious human being does, does not necessarily constitute it as being conscious. There is something more to consciousness. There is *something it is like* for us to be us. This is what Chalmers has coined as "the hard problem" of consciousness [13] which pertains to the problem of subjective experience – and more to the point, *why* we have such experiences. Thus, the hard problem makes it difficult to believe that we would be able to create artificial consciousness. How could we, if we do not even understand how our own phenomenal consciousness comes about, or if something like that does in fact exist (the jury is not yet out on the reductive/non-reductive physicalist debate)?

However, back to the focus of the paper, we may consider that given the capacity that social robots have to mimic consciousness, perhaps we need not be so overly concerned with the (im)possibility of robot consciousness. Thus, in the context of this paper, if it is the case that humans may interact with robots *as if* they are conscious, that is enough for us to argue that treating them immorally may negatively impact upon the moral fibre of interactants (and this is discussed in the section below).

As Arnold and Scheutz[3] state, it is "not what a robot is *in esse* but its function with and impact on people". The potential for interactants to perceive robots as being conscious stems from them being, as Turkle [58] states, a "relational artefact" in that these robots are "explicitly designed to engage a user in a relationship". This is due to their human-like appearance and social behaviour which work hand in hand to facilitate interactions that are as realistic as possible. Due to their human-like appearance and the capacity for robots to socially interact with interactants (albeit in a limited capacity[4]), there is a high possibility that interactants will anthropomorphise these robots. As Kanda et al. [32] state: a robot with a human-like body "causes people to behave unconsciously as if they were communicating with a human".[5]

However, in stating that interactants may anthropomorphise robots, this does not mean that they believe that these robots are *actually* human, but that does not mean human interactants cannot have relationships with robots. Rather, it means that a human-like appearance may evoke feelings within the interactant such that they view and treat their sexbots *as if* they are alive [37].[6] In the case of android social robots, if interactants want to perceive them as real people – as this may enhance their relational experience with them – then interactants may attribute human-like characteristics to them and treat them *as if* they are

[4] This is due to their incapacity to genuinely reciprocate human sentiments – related to the consciousness debate.

[5] The tendency to behave in such a way is brought about by the natural tendency that people have to anthropomorphise non-human entities or inanimate objects . Anthropomorphisation is an evolutionary trait inherent within us all (e.g. [17]).

[6] One can extrapolate that this may be the case from studies conducted with AIBO, a robotic dog, where Peter Kahn and his team stated: "We are not saying that AIBO owners believe literally that AIBO is alive, but rather that AIBO evokes feelings as if AIBO were alive" (see [37]).

human. This is no futuristic prediction. Studies have found that people do tend to apply social rules to the computers with which they interact [10] [45]. The more human-like something appears to be, the more likely we are to anthropomorphise it. As such, given their android appearance, it is no leap in logic to then argue that the tendency to anthropomorphise android social robots will likely be high.

Specifically in the context of social robots that may provide a form of companionship, it also may be the case that human interactants *want* to believe that the robot is conscious, because this will make their companionship with them seem all the more realistic (see e.g. [48] [6]), thus, human interactants may allow themselves to be deceived, thus *perceiving* the robots as conscious, although they may know that it is not *actually* conscious.

4.2 Treating social robots immorally does moral harm to human interactants

The human-like appearance of social robots, as well as their capacity to socially interact with us (albeit in a limited capacity) means, as was discussed above, that there is the possibility for human interactants to relate to social robots in a human-like way: we view them, and interact with them *as if* they are human beings, thus attributing to them human characteristics, such as consciousness. This relation is unique as compared to any other relation that we may have with other forms of technology. It is this unique relation that calls into question the morality of treating social robots immorally. I will here put forward that treating social robots immorally may morally harm human interactants. I argue this main point using three sub-arguments: social robots are more then mere objects; the act of treating a social robot immorally is abhorrent in itself; and treating a social robot immorally may negatively impact upon the moral fibre of interactants.

Social robots are more than mere objects. Although I am neutral for the purposes of this paper on whether or not social robots are capable of possessing consciousness – particularly in the phenomenological sense as has been discussed – I argue that we cannot deem them as merely being inanimate objects. We cannot place social robots within the same group as any other object we utilise. This is because we do not view and relate to social robots the same way in which we view and relate to any other objects in the world.

Dautenhahn [23], based on the work of Breazeal [8] [9], Fong et al. [26] and her own [22], elaborates upon how the definition and conceptual understanding of social robots may vary depending on their purpose and how and why they interact with people and the environment in which they are situated. Social robots can be: (1) "*Socially evocative*: Robots that rely on the human tendency to anthropomorphize and capitalize on feelings evoked when humans nurture, care [for] or [become involved] involve with their 'creation'" [8] [9], are socially evocative; (2) "*Socially situated*: Robots that are surrounded by a social environment which they perceive and react to [are socially situated]. Socially situated robots are able to distinguish between other social agents and various objects in the environment" [26]; (3) *"Sociable*: Robots that proactively engage with humans in order to satisfy internal social aims (drives, emotions, etc.) [are sociable robots]. These robots require deep models of social cognition" [8] [9]; (4) *"Socially intelligent*: Robots that show aspects of human-style social intelligence, based on possibly deep models of human cognition and social competence" [22], are socially intelligent; (5) *"Socially interactive*: Robots for which social interaction plays a key role in peer-to-peer HRI [Human-Robot Interaction], different from other robots that involve 'conventional' HRI, such as those used in teleoperation scenarios" [26], are socially interactive. Given these definitions and conceptual understandings of social robots, it is clear that social robots are a versatile technology, and that there are various ways in which human interactants can socially relate to them. As such, social robots cannot be compared to just any object that we utilize on a daily basis; we do not socially relate to just any inanimate object the way in which we may relate to a social robot.

Given that human interactants can socially relate to social robots, there is then the possibility for us to bond with them in seemingly realistic ways. Although any type of bond with a social robot may be unidirectional, and no type of reciprocation on the part of the robot truly indicates consciousness, the robot still does mimic reciprocation on a human social level, which impacts the humans with whom they interact. As such, I agree with Ramey [49] that there may be a unique social relationship (albeit possibly unidirectional as far as genuine reciprocation is concerned) between a human and a social robot that is qualitatively different from the way in which we relate to any other object that we utilise [49].

We have more than a physical relation to them. Yes, one can have more than a physical relation to an inanimate object – children, for example, love their stuffed toys and it can be argued that these toys are created to elicit an emotional response from children. However, this type of interaction and emotional response differs from that which we experience with social robots since stuffed toys do not reciprocate emotion, whereas social robots do – even though this reciprocity may be mere mimicry. Given this, interactants may begin to see social robots as being on the same plane as human beings (see e.g. [38]). Therefore, although they may not actually be conscious, we may view them as being such, given the human-like way in which we are able to relate to them (see e.g. [57] [31] [44]). Given this possibility, the superficial view to treat social robots as mere objects does not seem viable – there is more to them than that – although *actually* granting them consciousness and considering them deserving of moral treatment the way humans are, may be taking it a step too far, especially given the contentiousness of the consciousness debate (I will elaborate upon this point in a later section).

Given that I hold that social robots can be seen to be more than just any inanimate object due to the way in which we interact with them, I will now consider why the *act* of treating a social robot immorally is wrong in itself. This is because not only may social robots be viewed as being more than mere objects, but they can essentially be seen to be human simulacra in that that they are being designed in our image, so as to facilitate the possibility for us to have human-like relations with them.

The act of treating a social robot immorally is abhorrent in itself. Due to social robots being created to foster the possibility for people to potentially *view* them as being conscious and on the same plane as human beings, social robots may be said to ultimately be symbols of human beings. Therefore, any interaction with them is also symbolic of an interaction with a human being. Given this, one can argue that in treating a social robot immorally, one is *symbolically* treating a human immorally, and this act can be seen to be morally abhorrent in itself.

This may seem like a leap, but it is important to then home in (again) on the humanistic aspect of these robots. They are specifically designed and created so that interactants will easily anthropomorphise them and relate to them in a humanistic social capacity. This is the whole point of their creation – to be human simulacra in every possible way, both physically and behaviouristically. Studies have confirmed the potential for interactants to attribute human-like aspects to robots and treat them as if they are human (see e.g. [37] [32] [31]).

Social robots are designed and created so that when an interactant physically – and emotionally – interacts with them, they are essentially performing an act which simulates the act that would be performed with another human being. This is for instance why moral questions arise regarding whether it would be wrong to allow a human to play out a rape fantasy using a sex robot as the victim. Both Sparrow [52] and Turner [59] ask this question. I hold the view that such an act would be immoral because the human-like form of the robot is intended to be symbolic of a human being, and moreover, if there is the possibility that an interactant may behave unconsciously as if they are interacting with a human, then playing out a rape fantasy with a robot simulates the enactment of an immoral act upon a human being and this act is immoral in itself.

Therefore, the act of treating a social robot immorally is wrong in itself due to its symbolic meaning. If a human-like robot essentially symbolises a human being, and an interactant unconsciously behaves as if they are interacting with a human, yet treats this robot immorally, then the immoral act should be condemned. Due to the act itself being immoral, there may be subsequent negative implications that may arise if interactants do treat social robots immorally. It is therefore important to address not only the morality of the act itself, but also consider how the act of treating a social robot immorally may negatively impact interactants as moral beings.

Treating a social robot immorally may negatively impact upon the moral fibre of interactants. Given that we cannot deem social robots to be mere objects due to the way in which we view and relate to them, and due to the act of treating a social robot immorally essentially symbolising the act of treating a human immorally, there is the possibility that treating a social robot immorally may negatively impact upon the moral fibre of interactants. By this I mean that treating a social robot immorally may cause us to treat other humans immorally, similarly to the way in which Kant argues that the cruel treatment of animals may lead to us being “no less hardened towards men” [34].

“[T]o treat androids as humans is not to make androids actually human, but it is to make oneself an expanded self” [49] and the way we treat robots will affect ourselves and people around us. In light of this, Levy [38] argues that we should treat robots in the same moral way that we would treat any human because not doing so may negatively affect those people around us “by setting our own behaviour towards those robots as an example of how one should treat other human beings” [38].

Similar questions have been raised as far as the moral treatment of animals is concerned. Kant [33] makes the argument that we have the duty to ourselves to refrain from treating animals with violence or cruelty. This is because in treating animals immorally (with violence or cruelty) we “[dull] shared feelings of their suffering and so [weaken] and gradually [uproot] a natural predisposition that is very serviceable to morality in one’s relations with other men” [33]. Thus, immoral treatment of animals may negatively impact upon moral relation with other humans. Similarly, Turner [59] states: “If we treat animals with contempt, then we might start to do so with humans also. There is a link between the two because we perceive animals as having needs and sensations – even if they do not have the same sort of complex thought processes as we do. Essentially, animals exhibit features which resemble humans, and we are biologically programmed to feel empathy toward anything with those features”.

If there is concern raised about the way in which we treat animals extending to the way in which we treat humans, then surely there should be even more concern regarding our moral treatment of social robots which are realistic *human simulacra* as opposed to animals who may merely possess features that are exhibitive as human features? As such, going back to Levy [38], the main reason why he argues we should not treat robots immorally, is that if we take their embodiedness seriously, it would impact negatively on our social relations with humans if we treated them immorally. This argument stems from the possibility that there is the potential for people interact with social robots in seemingly realistic human-like ways, leading to the human interactant perceiving the robot as being sociable, intelligent and autonomous and, as such, being on the same plane as human beings. This being the case, if we do begin to perceive social robots as being on the same plane as human beings, Levy’s [38] argument that we should treat robots morally well, for our own sake, holds some weight.

One can, therefore, argue that since social robots are – in Levy’s [38] view – embodied computers, in treating a social robot immorally, one is simulating the immoral treatment of a human being (as I have discussed above). If we do come to view these robots as being on the same plane as human beings, and yet not respect them as human beings, one can question theoretically whether this will lead to desensitising us towards immoral behaviour, thereby lowering the moral barriers of immoral acts. Would this potentially lead to human beings treating one another in such immoral ways?

Although such an argument can be likened perhaps to similar ones, for instance, debates about the impact of violent video games or pornography on society, the argument about social robots differs in that “the nature of robots as three-dimensional entities capable of complex behaviours distinguishes them from other media” [52]. Therefore, treating a social robot immorally – by abusing them, or playing out a rape fantasy with them for instance – is more realistic than, say, video games, and, as such, is “more likely to encourage people to carry out the represented act in reality” [52]. As Turner [59] states: “[S]imulating immoral or illegal acts with robots harms human society in some way, by condoning or promoting an unpleasant behaviour trait: an instrumental justification. This is a similar justification to the reason why cartoons depicting child pornography are often banned – even though no child was directly harmed in the process”.

As such, I agree that our treating social robots immorally may negatively impact upon the moral fibre of human interactants.

4.3 Why human interactants are moral patients of their own agential moral actions towards social robots

Earlier sections discussed two instances wherein a human may be a moral patient in the context of their interaction with a robot. There is, however, also a third perspective that we can consider in this regard, and particularly in the context of social robots: a human interactant being a moral patient of their own agential moral actions towards a robot. In this instance, the human interactant would be the moral agent of their own actions, as well as the moral patient of those very same moral actions. That is, the impact of the very action taken by a human interactant towards a social robot is redirected towards the human agent, making them a patient of their own immoral actions because their moral fibre is impacted by the way in which the robot is treated (as was discussed above).

Specifically, in terms of moral patiency, I hold that human interactants are the *actual* moral patients, whereas robots as the *perceived* moral patients. The distinction between *actual* moral patiency and *perceived* moral patiency takes us back to my discussion on the *actuality* of robot consciousness and the *perception* of robot consciousness as there is an inextricable link between consciousness and moral patiency.

It is a commonly held belief that in order for something to have moral status, or be worthy of moral consideration, this something must be conscious in the phenomenological sense [7] [30], because this would mean that they are able to subjectively experience suffering; that they can *feel what it is like* to be a moral patient that is treated immorally. This type of consciousness refers to “the capacity for phenomenal experience or qualia, such as the capacity to feel pain and suffer” [7]. Thus, were we to consider the moral treatment of robots from a ‘robot perspective’, i.e. treating them well *for their own sakes,* this would imply that they can *actually* be moral patients in the sense that they can experience suffering at the hand of a human interactant who treats them immorally. This, in turn, would imply that robots are *actually* conscious. We saw, however, that the actuality of robot consciousness is a thorny issue and, therefore, I put forward that we focus our attention to *perceived* robot consciousness. Given the link between consciousness and moral patiency, we may consider that should human interactants *perceive* a social robot as being conscious, they may then *perceive* them as being moral patients; because a social robot can act *as if* they are conscious, they can therefore act *as if* they are suffering, should they be treated immorally.

Moral patiency can be understood as the case of being a target of moral action. In this instance, human interactants would not be *direct* targets of their own actions, but rather *indirect* targets – like a bullet ricocheting off its direct target and injuring an innocent bystander who becomes an indirect target of the shooter. They (human interactants) are indirectly impacted by way of their moral fibre being negatively impacted should they treat social robots immorally.

Where the robot is the direct target of the immoral treatment – and the *perceived* moral patient – the human interactant is the indirect target – and the *actual* moral patient. As such, we are indirect recipients of immoral action because robots cannot *actually* be recipients. Robots are not *really* impacted (for now leaving aside the possibility of robot phenomenal experience and consciousness, which, if it comes to pass, would of course add a layer of the robot as moral patient to this discussion) – we (the human interactants) are. Moreover, Danaher [19] notes a moral patient as "a being who possesses some moral status – i.e. is owed moral duties and obligations, and is capable of suffering moral harms and experiencing moral benefits – but who does not take ownership over the moral content of its own existence". As far as human interactants being moral patients of their own moral actions is concerned, referring to Danaher's [19] definition, human interactants can suffer and experience moral harms and benefits of their own agential actions: specifically, moral harms by way of their moral fibre being negatively impacted is an example of this kind of suffering.

Interestingly, Danaher [19] actually argues that the rise of robots could bring about a decrease in our own moral agency: "That is to say, [the rise of robots] could compromise both the ability and willingness of humans to act in the world as responsible moral agents, and consequently could reduce them to moral patients" [19]. For example, and as elaborated upon in Danaher's [19] article, an instance in which someone spends all their time with their sexbot. As a consequence, the human interactant loses motivation to do anything of real consequence – go out and meet new people, or spend time with a human partner – because it takes more effort. As such, this human interactant can spend all day at home, enjoying all the pleasure they desire [19]. As Danaher [19] states: "[T]he rise of the robots could lead to a decline in humans' willingness to express their moral agency (to make significant moral changes to the world around them). Because they have ready access to pleasure-providing robots, humans might become increasingly passive recipients of the benefits that technology bestows".

This is a compelling argument and worth consideration. However, I rather argue here not so much that our moral agency could itself 'decrease' due to our interaction with social robots but rather that our moral agency could be negatively impacted in the sense that as moral agents, our moral fibre may be negatively impacted, thus causing us, as moral agents, to possibly act immorally towards other human beings with whom we share the world, and towards ourselves.

Therefore, we may consider treating social robots morally well *for our own sakes*. Although specifically speaking to the topic of robot rights, we may here draw upon Gunkel's [28] argument that a consideration of the descriptive and normative aspects of robot rights seem to often be amiss in current machine ethics literature. It is important to distinguish between these two aspects so as to avoid slipping from one to the other. As far as the moral consideration of robots is concerned, this article distinguishes between the descriptive and normative aspects of the moral consideration of robos by way of arguing that even though social robots are not *capable* of being actual moral patients (descriptive aspect), we *should* still grant them moral consideration (normative aspect).

Finally, most ethics are agent-oriented – hence Floridi & Sanders [25] refer to this orientation as the 'standard' approach. As such, a patient-oriented approach is 'non-standard' – "it focuses attention not on the perpetrator of an act but on the victim or receiver of the action" [25]. Considering the possibility of human interactants being both agents and patients in a given instance bridges such a divide between a standard and non-standard approach. This is because human interactants – as moral agents – have the capacity to treat robots in moral or immoral ways. However, such treatment indirectly impacts human interactants as moral patients – they are, too, indirect receivers or victims of their own moral actions given that treating a robot immorally may negatively impact upon their own moral fibre.

5 Conclusion

This paper ultimately argued that given the *perception* of robot consciousness and moral patiency, as well as the possibility that treating a social robot immorally may cause moral harm to human interactants, we may consider that a human interactant is, at the same time, both a moral agent and a moral patient of their moral actions towards a social robot. That is, a human interactant (as a moral agent) is the *actual* moral patient of their moral actions, whereas the robots is a *perceived* moral patient.

This argument contributes to a perspective that is sorely lacking in machine ethics literature: there is very little focus on moral patiency as compared to moral agency (in the context of both humans and robots). Although there is somewhat of a focus on moral agency in that I argue that a human interactant is, at the same time, both the moral agent and the *actual* moral patient, there was more focus on human interactants being moral patients given that it it is more relevant in the context of an *anthropocentric* perspective on the moral treatment of robots. Moreover, a novel contribution is made particularly in the context of human moral patiency in the context of human-robot interaction. Where there has been consideration that humans can be moral patients in terms of robots being conduits of human moral action towards other human moral patients, as well as consideration that humans can be moral patients to the moral actions of robots, there has been no consideration of *human interactions being moral patients of their own agential moral actions towards robots* (particularly android social robots) i.e. indirect targets of their own moral actions, particularly in the context of treating robots immorally.

This is an important consideration and contribution in the context of the debate surrounding the moral treatment of robots, which also encompasses the contentious subject of robot rights. It is important because analysing the moral treatment of robots, and the possibility of robot rights, from an anthropocentric perspective (thus not in terms of whether or not robots are harmed from a robot perspective) as is suggested, may allow further research in this regard that does not become so concerned with the actuality of robot consciousness and moral patiency to such an extent that consideration concerning robot moral status and robot rights seem superfluous. The consideration of robot moral status and robots rights is definitely not superfluous from the perspective of human interactants who may be morally harmed as a result of immoral interactions with social robots who mimic human-likeness. The need to research the nature and impact of HRI is high and often under-estimated even in AI ethics policy making.

We cannot only consider the moral treatment of robots when, or if, they become conscious. The very way in which we express ourselves as humans and in which we situate ourselves in social spaces is in danger of changing rapidly already in the case of human traits simply being mimicked. To be detained by the concern as to whether robots can be conscious or not will only for now misdirect us from moral issues that should be immediately addressed and present more present ethical dangers: such as the degradation of our moral fibre due to not treating robots morally well for our own moral sakes.

As far as non-android social robots are concerned, further research may draw upon arguments I have made in the context of android social robots so as to possibly generalize arguments to the impacts of non-android social robots, or other types of robots in general. This, however, will require further research. Further research may also draw upon the arguments made so as to consider granting rights to robots. Specifically, we may consider granting negative rights to robots, i.e. rights that will prevent human interactants from treating robots immorally.

For now, the possibility of robots with full moral status who demand their rights may seem a long way off. We cannot be certain when this will happen, or if it will ever happen. Regardless of these possibilities, however, what we can be certain of is that the moral fibre of human societies may be at risk if we do not consider the moral treatment of social robots – at least, for now, from the perspective of human interactants.

References

1. IEEE ethically aligned design (2019), https://standards.ieee.org/content/dam/ieee standards/standards/web/documents/other/ead1e.pdf, online

2. Anderson, M., Anderson, S.: Machine ethics: Creating an ethical intelligent agent. AI Magazine **28**(4), 15–26 (2007)
3. Arnold, T., Scheutz, M.: HRI ethics and type-token ambiguity: what kind of robotic identity is most responsible? Ethics and Information Technology (2018)
4. Asaro, P.: What should we want from a robot ethic? International Review of Information Ethics **6**(12), 9–16 (2006)
5. Barquin, R.C.: Ten commandments of computer ethics (1992)
6. Boltuć, P.: Chuch-Turing Lovers. s.l.:Oxford University Press (2017)
7. Bostrom, N., Yudkwosky, E.: The ethics of artificial intelligence. In: Frankish, K., Ramsey, W. (eds.) The Cambdridge Handook of Artificial Intelligence, p. 316–334. Cambridge University Press, Cambridge (2014)
8. Breazeal, C.: Designing sociable robots. MIT Press, Cambridge, MA (2002)
9. Breazeal, C.: Towards sociable robots. Robotics and Autonomous Sys-tems **42**, 167–175 (2003)
10. Broadbent, E.: Interactions with robots: The truths we reveal about ourselves. Annual Review of Psychology **68**, 627–652 (2017)
11. Brundage, M.: Limitations and risks of machine ethics. Journal of Experimental & Theoretical Artificial Intelligence **26**(3), 355–372 (2014)
12. Bryson, J.: Robots should be slaves. close engagements with artificial companions: Key social, psychological, ethical and design issues (2009)
13. Chalmers, D.: Facing up to the problem of consciousness (1995), http://consc.net/papers/facing.pdf, [Accessed 7 May 2019]
14. Chalmers, D.: Philosophy of Mind: Classical and Contemporary Readings. Oxford University Press (2002)
15. Coeckelbergh, M.: Artificial companions: Empathy and vulnerability mirroring in human-robot relations. Studies in Ethics Law and Technology **4**(3, Article 2) (2010)
16. Coeckelbergh, M.: Health care, capabilities, and ai assistive technologies. Ethical Theory and Moral Practice **13**, 181–190 (2010)
17. Damiano, L., Dumouchel, P.: Anthropomorphism in human–robot co-evolution. Frontiers in Psychology **9**, 1–9 (2018)
18. Danaher, J.: The Symbolic-Consequences Argument in the Sex Robot Debate. MIT Press, Cambridge (2017)
19. Danaher, J.: The rise of the robots and the crisis of moral patiency. AI & Society **34**, 129–136 (2019)
20. Danaher, J., Earp, B., Sandberg, A.: Should We Campaign Against Sex Robots? The MIT Press, Cambridge, MA (2017)
21. Danaher, J., McArthur, N.: Robot Sex: Social and Ethical Implications. The MIT Press, Massachusetts (2017)
22. Dautenhahn, K.: The art of designing socially intelligent agents - science, fiction, and the human in the loop. Applied Artificial Intelligence **12**, 573–617 (1998)
23. Dautenhahn, K.: Socially intelligent robots: dimensions of human–robot interaction. Philosophical Transactions of the Royal Society p. 679–704 (2007)
24. Deng, B.: Machine ethics: the robot's dilemma. Nature News **523**, 24–26 (2015)
25. Floridi, L., Sanders, J.: On the morality of artificial agents. Mind and Machines **14**, 349–379 (2004)
26. Fong, T., Nourbakhsh, I., Dautenhahn, K.: A survey of socially inter-active robots'. Robots and Autonomous Systems **42**, 143–166 (2003)
27. Gunkel, D.: Moral patiency. In: The Machine Question: Critical Perspectives on AI, Robots, and Ethics, p. 93–157. The MIT Press, Cambridge, MA (2012)
28. Gunkel, D.: The other question: Can and should robots have rights?'. Ethics and Information Technology **20**, 87–99 (2017)

29. ICRC: Autonomy, artificial intelligence and robotics: Technical aspects of human control. s.n, Geneva (2019)
30. Jaworska, A., Tannenbaum, J.: The grounds of moral status. Stanford Encyclopedia of Philosophy (2018)
31. Kanda, T., Freier, N., Severson, R., Gill, B.: Robovie, you'll have to go into the closet now": Children's social and moral relationships with a humanoid robot. Developmental Psychology **48**(2), 303–314 (2012)
32. Kanda, T., Ishiguro, H., Imai, M., Ono, T.: Development and evaluation of interactive humaoid robots. s.l pp. , 1839–1850 (2004)
33. Kant, I.: The Metaphysics of Morals. Cambridge University Press, Cambridge (1996)
34. Kant, I.: Lectures on Ethics. Cambridge University Press, Cambridge (1997)
35. Kirk, R., Carruthers, P.: Consciousness and concepts. In: Proceedings of the Aristotelian Society, Supplementary. vol. Volumes, Volume 66, p. 23–59 (1992)
36. Komatsubara, T.: Can a social robot help children's understanding of science in classrooms? In: s.l., Proceedings of the second international conference on human–agent interaction. p. 83–90 (2014)
37. Levy, D.: Love and sex with robots: The evolution of human-robot relationships (2007), s.l.:Harper.
38. Levy, D.: The ethical treatment of artificially conscious robots. International Journal of Social Robotics **1**(3), 209–216 (2009)
39. Lin, P., Abney, K., Bekey, G.: Robot Ethics: The Ethical and Social Implications of Robotics. The MIT Press, Cambridge, MA (2012)
40. Lin, P., Abney, K., Bekey, G.: Robotics, Ethical Theory, and Metaethics: A Guide for the Perplexed'. The MIT Press, Cambridge, MA (2012)
41. Lumbreras, S.: The limits of machine ethics. Religions **8**(100), 2–10 (2017)
42. Matthias, A.: The responsibility gap: Ascribing responsibility for the actions of learning automata. Ethics and Information Technology **6**, 175–183 (2004)
43. McDermott, D.: Why Ethics is a High Hurdle for AI. North American Conference on Computers and Philosophy, Bloomington, Indiana (2008)
44. Melson, G., Kahn, P., Beck, A., Friedman, B.: Robotic pets in human lives: Implications for the human–animal bond and for human relationships with personified technologies. Journal of Social Issues **65**(3), 545–567 (2009)
45. Moon, Y., Nass, C.: Machines and mindlessness: social responses to computers. Journal of Social Issues **56**, 81–103 (2000)
46. Moor, J.: The nature, importance, and difficulty of machine ethics. IEEE **21**(4), 18–21 (2006)
47. Müller, V.: Ethics of artificial intelligence and robotics (2020), edition).
48. Nyholm, S., Frank, L.: It loves me, it loves me not: Is it morally problematic to design sex robots that appear to love their owners? (2019), techné: Research in Philosophy and Technology, Issue December.
49. Ramey, C.: 'For the sake of others': The 'personal' ethics of human-android interaction. Stresa, Italy (2005)
50. Sharkey, A., Sharkey, N.: Granny and the robots: Ethical issues in robot care for the elderly. Ethics and Information Technology **14**(1), 27–40 (2010)
51. Sharkey, A.: Should we welcome robot teachers? Ethics and Information Technology **283–297**, 18 (2016)
52. Sparrow, R.: Robots, rape, and representation. International Journal for Social Robotics **9**(3), 465–477 (2017)
53. Sparrow, R., Sparrow, L.: In the hands of machines? the future of aged care. Minds and Machines **16**, 141–161 (2006)

54. Sullins, J.: When is a robot a moral agent? International Review of Information Ethics **6**(12) (2006)
55. Sullins, J.: Robots, love and sex: The ethics of building a love machine. IEEE Transactions on Affective Computing **3**(4), 398–409 (2012)
56. Torrance, S.: Artificial agents and the expanding ethical circle. AI & Society **28**, 399–414 (2013)
57. Turkle, S.: A Nascent Robotics Culture: New Complicities for Companionship. AAAI Technical Report Series (2006)
58. Turkle, S.: Authenticity in the age of digital companions. Interaction Stud-ies **8**(3), 501–517 (2007)
59. Turner, J.: Why robot rights? In: Robot Rules: Regulating Artificial Intelligence, p. 145–171. Palgrave Macmillan, Cham (2019)
60. Vallor, S.: Carebots and caregivers: Sustaining the ethical ideal of care in the twenty-first century. Philosophy & Technology **24**, 251–268 (2011)
61. Wallach, W., Allen, C.: Moral Machines: Teachinf Robots Right from Wrong. Oxford University Press, New York (2009)
62. Wang, W., Siau, K.: Ethical and Moral Issues with AI: A Case Study on Healthcare Robots. Twenty-fourth Americas Conference on Infor-mation Systems, New Orleans (2018)