Social robots and the risks to reciprocity

Aimee van Wynsberghe¹

Received: 24 April 2020 / Accepted: 25 March 2021 / Published online: 30 April 2021
© The Author(s) 2021

Abstract

A growing body of research can be found in which roboticists are designing for reciprocity as a key construct for successful human–robot interaction (HRI). Given the centrality of reciprocity as a component for our moral lives (for moral development and maintaining the just society), this paper confronts the possibility of what things would look like if the benchmark to achieve perceived reciprocity were accomplished. Through an analysis of the value of reciprocity from the care ethics tradition the richness of reciprocity as an inherent value is revealed: on the micro-level, as mutual care for immediate care givers, and on the macro-level, as foundational for a just society. Taking this understanding of reciprocity into consideration, it becomes clear that HRI cannot achieve this bidirectional value of reciprocity; a robot must deceive users into believing it is capable of reciprocating to humans or is deserving of reciprocation from humans. Moreover, on the macro-level, designing social robots for reciprocity threatens the ability and willingness to reciprocate to human care workers across society. Because of these concerns, I suggest re-thinking the goals of reciprocity in social robotics. Designing for reciprocity in social robotics should be dedicated to the design of robots to enhance the ability to mutually care for those that provide us with care, as opposed to designing for reciprocity between human and robot.

Keywords Social robots · Robot ethics · Care ethics · Reciprocity · Human–robot interaction

1 Introduction

Social robots, or robots *perceived* to be social, have gone from academic curiosity to commercial products used in hospitals, care facilities, schools, and homes. Some roboticists are pursuing the development of social robots to encourage acceptance and ease of use of robots (Lammer et al. 2014) whilst others aim to use robots to ‘nurture or nudge pro-social behaviours’ amongst humans (Paiva et al. 2018). Of late, a growing body of research can be found in which roboticists identify the value of reciprocity as a “key mechanism of human interaction” (Lorenz et al. 2016, p. 1) to be explicitly included in the design of future social (companion) robots in therapy and care (Fischinger et al. 2016; Kahn et al. 2006; Lammer et al. 2014; Lorenz et al. 2016; Sandoval et al. 2016). In general, the idea is that if humans and robots engage in an interaction based on mutual

care, aka reciprocity, humans will perceive such interactions positively (e.g. the robot is easier to use) and will accept the robot. At first glance this may seem convincing but upon closer inspection, we ought to question the impact of nudging reciprocal interactions between humans and robots given the significance of reciprocity as a value within a just society.

Across a range of disciplines reciprocity is thought to be “a central feature of moral life” (Kahn et al. 2006, p. 368). It is a component of moral development (Duska and Whelan 1975), instrumental to good care and an ‘indispensable duty’ for the just society (Gouldner 1960; Tronto 1993). With the increase in both the availability of robots perceived to be social, and the intention to design for reciprocity, comes the need for ‘moral learning’ (Van De Poel 2018) about the ethical impact resulting from these new kinds of reciprocal interactions between humans and robots. To be sure, robot ethicists have studied the relationship between robot design and societal values for some years now (i.e. ‘how to embed values in robot design’ and/or ‘how to understand the impact of a robot’s design on cultural values’) (Arnold Scheutz 2017; Calo, 2011; Decker 2008; Sharkey 2014, 2014; van Wynsberghe 2013a, b); however, never has the

✉ Aimee van Wynsberghe
aimee@uni-bonn.de

¹ Alexander Von Humboldt-Professor for Applied Ethics of Artificial Intelligence, University of Bonn, Bonn, Germany

value of reciprocity been on the list of values to address. It is my aim in this paper to explore the meaning of reciprocity through the lens of care ethics as a step towards understanding the significance of this value as more than a benchmark for designers, but as a construct central to good care and the just society.

In the following paper, I begin by briefly outlining the motivations and assumptions underpinning the movement to design social robots for reciprocity. I suggest that given the growing number of experiments (and literature) dedicated to the embedding of reciprocity in human–robot interactions (HRI) that it is time for robot ethicists to question: what happens if the aim of creating reciprocal HRI goes right and what happens if it goes wrong. In so doing I suggest a specific kind of deception occurring between social robots and human users; namely, humans are deceived into believing that the robot is deserving of reciprocity by the robot's appearance of responsiveness. Following this, I address the need to look beyond dyadic HRIs and to assess the impact of designing for reciprocity on the macro-level. As robots enter into social care practices, there will be a re-organization of roles, responsibilities and resources (van Wynsberghe A Li 2019) and social robots may threaten the ability to reciprocate to, and further may weaken the incentive to give back to, care workers. Based on the above discussion, I suggest roboticists re-think the aims of social robotics at large: the pursuit of social robots should be to enhance human–human reciprocity rather than human–robot reciprocity.

2 Designing social robots for reciprocity

One of the earlier definitions of social robots, articulated by Breazeal, defines them as robots “designed to interact with people in a socio-emotional way during interpersonal interaction” (Breazeal 2004). Others have suggested that the field of social robots concerns itself with “how the robot should interact with humans in order to be perceived as social” (Frennert Östlund 2014). The latter, broader, definition opens up the possibility that any kind of robot, e.g. service, industrial, may have social capabilities ‘in order to be *perceived* as social’ rather than be called social robots per say. Take for example robots to greet people in banks and hospitals, the goal is not necessarily to establish a long-term social relationship; however, the goal is to interact in such a way that the robot is *perceived* to be social and thus humans are more likely to accept and engage with the robot. This paper concerns itself with the category of robots designed to be perceived as social rather than the more rigid definition of social robots (as stated above). This implies that a factory robot, such as Baxter, that is perceived to be social could fall within the scope of my critique here.

To create the perception of being social, or to create a lasting relationship between human and robot, the HRI is almost always based on the characteristics of human–human interactions (HHI) (Kahn et al. 2006; Sandoval et al. 2016). Given that HHIs form the baseline for creating HRIs, (social) roboticists have explored the variety of features found in HHI that ought to be embedded into the HRI so as to ensure intuitive and successful interactions. In fact, Kahn et al. created a list of six benchmarks present in HHI that can be used to assess the success of HRIs. Amongst the elements on this list is reciprocity. Reciprocity can be simply defined as the “Golden Rule”: do unto others as you would have them do unto you (Kahn et al. 2006). Or, “If you do something for me I will do something for you” (Sandoval et al. 2016). In any case, the idea is to treat others in a manner reflective of how they have treated you, either positively or negatively. The idea of the benchmark then is to ask whether people can “engage substantively in reciprocal relationships with humanoids” (Kahn et al. 2006, p. 268).

Since that time roboticists have continued to explore the embedding of reciprocity into HRIs with the assumption being that reciprocity plays a vital role in HHI and “it is expected that reciprocity plays an important role in HRI” (Sandoval et al. 2016, p. 304). Accordingly, Fogg and Nass “demonstrated that users provided more helping behaviour to a computer that had helped them previously than to a different computer. Users also worked longer, performed higher quality work, and felt happier” (Fogg Nass 1997, p. 331).

In 2013, a workshop was held as part of the International Conference of Social Robotics in which the importance of reciprocity for social robotics was explicitly discussed. The workshop entitled “Taking care of Each Other: Synchronization and Reciprocity for Social Companion Robots” explored a variety of concepts under the umbrella of reciprocity thought to serve as the “cornerstone in the future development of meaningful HRIs” (Sandoval et al. 2016, p. 303). Accordingly, it is believed that “in the future, robots could assume more social roles in the human domain if the HRI would be more reciprocal” (Sandoval et al. 2016, p. 304).

Experimentation along these lines has been conducted involving children with an AIBO robot showing that “children responded reciprocally and were more engaged with an AIBO robot which offered some motioning, behavioural and verbal stimulus than they were with the toy dog” (Kahn et al. 2004). Other studies have been conducted with elderly persons using the concept of reciprocity as “mutual care” demonstrating that “reciprocal behaviour, even in short term laboratory studies, positively influences the perceived usability and ease of learning of the care robot” (Lammer et al. 2014). In the first set of studies—involving children—the goal was to investigate whether children would reciprocate to the robot whereas in the second set of experiments—involving older adults—the goal was to display the robot

as reciprocating towards the human user so as to increase acceptability of the robot. Each used reciprocity as a design paradigm; however, they used the construct differently (i.e. for the human to reciprocate to the robot or for the robot to reciprocate to the human).

With the increase in both the availability of robots perceived to be social, and the intention to design for reciprocity, comes the need for ‘moral learning’ (Van De Poel 2018) about the ethical impact resulting from these new kinds of reciprocal interactions between humans and robots. What could go right if roboticists meet the benchmark for reciprocity? Is fostering reciprocity in HRI desirable on a broader scale, in the long-term? Although the concept of reciprocity has been studied in a variety of disciplines, e.g. moral development (Duska Whelan 1975) and sociology (Gouldner 1960) to name a few, reciprocity has (historically) rarely been on the list of values to embed in robotic systems and thus rarely caught the attention of robot ethicists. To address these questions I suggest turning our attention to the care ethics domain to shed light on the definition, significance and meaning of reciprocity in the kinds of interactions that constitute HRIs with social robots, namely interactions of care.

3 Care ethics and reciprocity

Care ethics as a moral tradition matured out of a growing need in the 1980s to shift the focus of moral theory from Kantian rule based theories and/or consequentialist perspectives towards moral reasoning that places relationality at the core. Care ethicists Carol Gilligan, Nel Noddings, and Joan Tronto are amongst the most notable scholars working to articulate the: meaning of care as a combination of attitude and action; understanding of ethical concepts such as autonomy as relational versus atomistic; and understanding care as a political construct which demands that care serves as a basis for ‘political practice’ (Tronto 1993). Together these scholars have paved the way to make care a formidable contribution to the ethical and political discourse of the twenty-first century.

Common attributes of an ethic of care hold that good care is first and foremost relational. Further, good care is manifest through both actions and intentions. These actions cannot be reduced to isolated events or tasks; rather, they ought to be understood as complex practices in which numerous actors are involved in the identification of needs, the meeting of needs, and the response to needs. These complex processes are labelled by Tronto as care practices for which there are four iterative steps: (1) *Caring about*; someone or some group notices unmet caring needs; (2) *Taking care of*; someone or some group has to take responsibility to make certain that these needs are met; (3) *Caring-giving*; actual

care-giving work is done; (4) *Care-receiving*; throughout the provision of care work the response from the person, thing, group, animal, plant, or environment is observed and together judgements are made about it (Tronto 1993, pp. 105–108).

This last step can be assessed, according to Tronto through the moral element of *responsiveness* which essentially entails that care givers “consider the other’s position as that other expresses it” (Tronto 1993, p. 136). As such, good care is assessed through the care receiver’s positive change in functioning and through the care receivers perception of the care provided.

The element of responsiveness re-calibrates care from a unidirectional action to a bidirectional practice in which care givers are both attentive to the particular needs of care receivers as well as being attentive to if, when, and how such needs are met. In short, care should not happen *to* someone (or something); rather, care happens *with* someone (or something). This understanding is meant to cast-off traditional conceptions of care receivers as passive accepters of care. Instead, care receivers too have a role to play in the care process—they have opinions, preferences, desires, and hopes; they are holistic persons with life histories that should be shared as part of good care (Vanlaere Gastmans 2011). Recognizing this is a step towards amending the asymmetry in power between the powerful care giver and the ‘vulnerable’ care receiver (Verkerk, 2001).

This bidirectional conception of care aligns with another construct in care ethics; namely, reciprocity. Reciprocity and responsiveness share similarities: they both aim at empowering the otherwise vulnerable stakeholders in care; and they both illustrate that care is not one-sided.

To be sure, reciprocity is not the same as responsiveness: reciprocity is about giving back to others when they have done something for us, for example, when a person grants me a favour I want to return the favour in kind. Alternatively, responsiveness is about reacting to the actions of another, e.g. saying ‘this position makes me sore, can you help me move to another position’.

It is important to note that reciprocity as a moral construct in the care ethics tradition has been described on both a dyadic level and on a political level. On a dyadic, or micro-level, it structures our commitment to immediate care givers whereas on a political, or macro-level, it structures our commitments to care givers at large. Reciprocity on a macro-level is about mutual care for those in society who serve as care givers. Tronto insists that everyone is in need of care to greater and lesser extents at different moments of their lives and as such society should do away with the myth that to be in need of care is to be incomplete. Instead, if we recognize the ubiquity and universality of care then we may come closer to appreciating those who provide care. In line with this thinking, Tronto extends the thinking of reciprocity

beyond a care relationship and insists it must also be understood as a political ideal that focuses on giving back to care givers in a broader sense, e.g. making sure care givers are not economically punished for their role as care givers.

In essence, reciprocity in this broader sense should be understood “both as a quality of the just state and as a principle for generating the obligation to support cooperative caring schemes that provide care well, and in a just manner” (Sander-Staudt 2015, p. 196). From this vantage point, reciprocity is about recognizing the care work being done across society and to insist on ‘broad social responsibility’ to return care to those who have provided it; “creating social institutions that enable care-givers to do the job of caretaking without becoming disadvantaged in the competition for social benefits” (Kittay 109).

From the lens of care ethics, the value of reciprocity is inherently valuable, it is a good to aim for on its own and not necessarily for attaining some other end. On the micro-level, it is a value that when expressed restores power to the otherwise vulnerable cohorts. And on the macro-level, mutual care for care givers is a pillar of the just society. Understanding reciprocity as a concept rich in meaning then we ask, what does it mean to design for reciprocal interactions in HRI?

3.1 What happens if things go right: meeting the benchmark for reciprocity

As mentioned earlier, in a paper by Kahn et al., an idea was proposed to create psychological benchmarks for successful HRIs, to measure “categories of interaction that capture conceptually fundamental aspects of human life...” (Kahn et al. 2006, p. 364). In addition, reciprocity is one of these six benchmarks. The benchmarks are psychological (as opposed to ontological) in so far as the authors are suggesting that the robot need not ontologically be understood as having the capability of reciprocity but must be psychologically *believed* or *perceived* to be capable of reciprocity. So what happens if this benchmark is achieved; what happens if things go right?

In the “Mutual Care” project, in which the Hobbit robot was used, older adults were assessed to understand how they would react when engaged in an HRI that involved mutual care, i.e. when the human should help his/her care robot. The mutual care paradigm created HRI scenarios using the ‘rule of reciprocity’, that “we should try to repay, in kind, what another person has provided us”. The result was that the human participants did in fact believe that the robot and human supported one another. Consequently, the researchers conclude that in fact they had achieved *perceived reciprocity* (Lammer et al. 2014, p. 6), that is, the participants believed the robot to be capable of reciprocal interactions in which both the robot and the human were in need. Is there a harm

in deceiving people into thinking a robot can reciprocate kindness to humans?

Amongst the current academic literature, there is a certain spectrum of ethical issues related to the tendency for social robots to deceive users. Sharkey and Sharkey raise ‘deception and infantilisation’ as one of the top six ethical concerns attributed to social robots (Sharkey Sharkey 2012). For Sullins, robotic companions can only meet one’s physical and emotional needs on the surface but a robot cannot truly satisfy these needs even if the human is deceived into thinking the robot can (Sullins 2012). For Scheutz, the deception of humans by social robots is characterized as a ‘unidirectional bond’ between human and robot; the robot cannot bond in the same way a human does even if the human thinks it can (e.g. by projecting intentions onto the robot) (Scheutz 2012). This risk of this unidirectional bond, for Scheutz, is that companies will be able to exploit it for commercial gains. In short, for many robot ethics scholars of today, there appears to be something disturbing about unidirectional social relationships especially when such a starting point is used to steer the design and development of social robots.

Contrary to these ideas, other scholars have suggested that “As humans, most of the time we play a certain role to some extent and it is difficult to distinguish between what appears to be and reality in human social life” (de Graaf 2016). The author continues to suggest that “putting all emphasis on deception when evaluating human–robot relationships seems misplaced. As long as the user perceives to be served well by a robot and is satisfied with the behaviour of that robot, there should not be a problem in this account” (de Graaf 2016). Such a conclusion is dangerous as it fails to acknowledge that deception is not something that should be acceptable in the technologies that we humans create. It fosters a kind of naturalistic fallacy to suggest that the negative realities of HHI *ought* to be translated into HRI. Furthermore, it fails to account for the plurality of ways in which deception can be understood and more importantly, disregards the seriousness of the risk of deception to the foundational elements of social relationships.

I suggest that rather than disregard deception as a threat to social robots because people can deceive other people, the focus should instead be on delineating a particular kind of deception: social robots designed for reciprocity use reciprocity as an instrumental value to enhance acceptability of the robot. Further, the robots are meant to deceive users about the robot’s ability to engage in reciprocal relationships; when a robot is responding (to a command) it can appear to be an act of reciprocation. Furthermore, such ‘faux’ acts of reciprocity call upon humans to reciprocate to the robot.

Take for example the earlier studies done with the vacuum cleaning robot Roomba. The robot is not a social robot per se but studies have now shown that humans project

their own beliefs about the robot's social maturity onto the robot, thus the robot is perceived to be social. This was done to such an extent that after extensive experience with the robot in which the robot was meeting the needs of users (e.g. keeping the floor clean), human users began to treat it as an entity deserving of reciprocity. "The mere fact that an autonomous machine keeps working for them day in day out seems to evoke a sense of, if not urge for, reciprocation. Roomba owners seem to want to do something nice for their Roombas even though the robot does not even know that it has owners" (Scheutz 2012, p. 7). We can see here that it is not only the reality of unidirectional bonds (and the consequences of exploitation sketched by Scheutz) but that the robot is able to draw the human in to a situation in which the human is deceived into thinking he/she ought to reciprocate *to the robot* the good that the robot has provided them rather than understanding that the robot is merely responding to the command it has been given and/or the role it has been assigned.

Even still, aside from the fact that humans are being intentionally deceived into believing that the robot is capable of engaging in reciprocal interactions, is it possible to suggest that some good could come from this? Another area of research has proposed autonomous systems (aka robots) for promoting pro-social behaviour. The goal in this space is to create robots to "nudge or nurture cooperation and pro-sociality in a society of humans and machines" (Paiva et al. 2018). For this group of researchers, the starting point is the 'empathy deficit', a general lack of empathy development in younger generations. The idea then is to create a machine that can help to establish or foster cooperation, and even empathy, through the robotic interface or the robot as part of a team of actors.

Setting aside concerns of technological solutionism, if we take this thinking to its logical conclusion, then humans would be able to interact with robots so as to learn the skills needed to be empathic, cooperative, and reciprocating members of society. We should be clear, however, that this is not the aim for social robots designed for reciprocity with older adults in their own homes—the robots are not instruments to train older adults to have these moral skills. Instead, the aim is to encourage "acceptance towards the robot" (Lammer et al. 2014, p. 2). Alternatively, 'learned reciprocity' may very well be the aim when using social robots in experiments involving children. But an even larger problem exists here: how can one be sure that the skills learned through HRI will transfer to HHI?

To start, the 'needs of the robot' for which a human ought to reciprocate are recharging the battery, updating software, etc. Reciprocity towards the robot is not the same as reciprocity towards a human. Moreover, the paradigm within which the HRI exists is with an obedient, subservient agent (the robot) and this is not the HHI paradigm (ideally). If

humans have been trained for reciprocation with a subservient robot will their skills transfer over to an unpredictable HHI in which a human interaction partner does not take on such a subservient role? Granted these questions are entirely empirical, they are questions for which measurements will need to be created to understand if reciprocity was causally related between HRI and HHI, measurements that will also have to take into account the variety of ways in which reciprocity is understood and achieved between cultures and groups of people. The incredible amount of resources (time and money) needed to prove that the goal of pro-sociality, and reciprocity in particular, is achievable cannot be justified in a world with finite resources.

Consequently, if researchers get it right and are able to design for reciprocity in an effort to increase the chances of users accepting the social robot, they do so by deceiving users into believing that a social robot is capable of reciprocity or is deserving of reciprocity. One may make the justification that this deception is acceptable given that reciprocity is a fundamental component of moral development and as such HRI could be instrumental in fostering reciprocity in HHI; however, this hypothesis rests on the belief that the skills developed during HRIs would be transferable to HHIs. All of these considerations aside, it is also necessary to acknowledge that these are reflections on the HRI as a dyadic interaction between human and robot and they pay little to no attention to the reality that robots are entering complex sociotechnical systems and further that reciprocity also holds value and meaning on a broader societal level.

3.2 Reciprocity at the macro-level

Studies in HRI to date have focused on a dyadic interaction between human and robot and in so doing, these studies fail to account for the reality that robots are never just interacting with one human but are instead interacting with a complex system of actors (both human and non-human) (van Wynsberghe Li 2019). As such, the wider social context must be taken into consideration when assessing or evaluating social robots given that the social robot will bring a 'redistribution of tasks and responsibilities' throughout the system (van Oost Reed 2011).

Take healthcare as an example, "one of the most critical aspects of introducing robots in healthcare is how such a 'bot' will re-structure the healthcare system in a variety of ways: roles of healthcare staff will change once 'bots' are delegated tasks, certain professions may no longer exist (e.g., cleaning robots may remove the need for janitorial staff), the education of healthcare staff will need to include 'bot' training, resources will be re-allocated to account for the purchasing of 'bots', the expertise of healthcare staff will be called into question (e.g., when an AI algorithm predicts something that the physician doesn't)" (van Wynsberghe Li

2019). In other words, the introduction of a social robot into any care practice, and into society in general, is the introduction of a social robot into a sociotechnical system. Building on this it is necessary to consider what resources (e.g. time, money) will be taken away from human care givers and re-directed to the social robot?

If social robots are a replacement for human care givers (as they will be when living with older adults in their homes), then finite resources such as time and money will be directed towards the social robot rather than towards human care givers. Money will be invested in the production, implementation and maintenance of the robot rather than in the hiring and training of new care workers. Education of care workers (occupational therapists who visit the homes of people to provide care) will change to include ‘how to interact with’ or manage the social robot in the home. Expertise of the care workers will be called into question if the robot provides different advice or recommendations to the human care worker. In essence, the ability to reciprocate will be minimized as the finite resources are directed towards the integration of social robots into care practices. Moreover, the willingness or interest in reciprocating to human care workers may suffer as a result of the robot’s ability to draw human users in to believe that it requires reciprocity. If a Roomba robot can convince owners it is deserving of a ‘break from cleaning’, it is possible to consider that a hospital delivery robot may convince care workers, it is deserving of a ‘break from delivering’. The hospital care workers may be further burdened by the robot, whose implementation was intended to relieve a burden, when reciprocity is directed towards the robot rather than the human care workers.

4 Conclusion: re-thinking reciprocity for social robotics

In the current global pandemic of 2020, one will find a series of news articles showing a lack of mutual care for the care givers that we all rely on right now to save us. At the same time, one can also find an increase in the interest to design for reciprocity in interactions between humans and robots. This paper was meant to confront the possibility of what things would look like if the benchmark to achieve perceived reciprocity were accomplished—an exploration of the lessons we should be learning when we take a closer look at the empirical work happening with the design of HRI for reciprocity. Through an analysis of the construct of reciprocity from the care ethics tradition the richness of reciprocity as an inherent value is revealed: on the micro-level, as mutual care for immediate care givers, and on the macro-level, as foundational for a just society. Taking this understanding of reciprocity into consideration, it becomes clear that HRI cannot achieve this bidirectional value of reciprocity; robots

must deceive users into believing it is capable of reciprocating or is deserving of reciprocation. Moreover, on the macro-level, designing social robots for reciprocity threatens the ability and willingness to reciprocate to human care workers across society.

Because of these concerns, I suggest re-thinking the goals of reciprocity in social robotics. Designing for reciprocity in social robotics should be focused on the design of robots to enhance the ability to mutually care for those who provide care. Take for instance, exoskeleton robots worn by nurses to relieve the heavy lifting of patients (by nurses) whilst the nurse carries out his daily tasks. Or, telecommunication robots (like the RP-7) that can be used to keep healthcare workers at a safe distance from patients during a pandemic and yet still able to communicate with patients. Each of these robots is meant to provide care for those who are in the midst of providing care—they are robots that manifest the value of reciprocity between humans by the very use of the robot.

To achieve this way of thinking there needs to be a concerted effort towards the ‘design for reciprocity’ amongst humans. For robot designers, one might consider following a care-centred value sensitive design approach (van Wynsberghe 2013a, b; 2016) with reciprocity as the foundational value to be embedded in the HRI. Furthermore, as this paper argues, robot designers, when designing for reciprocity, ought to conceptualize it as a value to be established between humans rather than as a value to be established between the human and the robot.¹

Equally important, reciprocity need not be understood as a value exclusive to the healthcare context. There are a variety of spaces in which mutual care, reciprocity, is necessary and yet sorely lacking. Consider the problem of electronic waste (e-waste) and the lack of mutual care towards the communities (at a distance) where the majority of e-waste is shipped and stored. These communities bear the burden of society’s demand for electronic devices and yet there is little done to care for the growing needs they have as a result of this e-waste (e.g. the minerals from e-waste pollute the soil and water affecting the wellbeing of these communities, thus there is a need for clean soil and clean water). Social robots in this space could be about the design of robots to separate and clean e-waste thereby preventing humans from exposure to the chemicals.

Robotics in general should be about providing support for those who support us in our lives. Social robots in particular should not be about creating faux reciprocal relationships between humans and robots, that is, relationships that are deceptive and unidirectional at their core. To create robots

¹ Consider the value of safety as an example; when designing robots to be safe the value of safety is understood as a consequence of the human and robot interacting.

with the intention to deceive users threatens the value of reciprocity across society. In addition, a world without reciprocity, without mutual care, is a world entirely unsustainable.

Acknowledgements This research is supported by the Netherlands Organization for Scientific Research (NWO), project number 275-20-054. Many thanks to Johanna Seibt and Marco Nørskov for inviting me to present an earlier version of this work at the 2020 RoboPhil conference.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Arnold T, Scheutz M (2017) The tactile ethics of soft robotics: designing wisely for human-robot interaction. *Soft Robot* 4(2):2. <https://doi.org/10.1089/soro.2017.0032>
- Breazeal CL (2004) Designing sociable robots. MIT Press
- Calo R (2011) Robots and privacy. In: Lin P, Abney K, Bekey GA (eds) *Robot ethics: the ethical and social implications of robotics*. MIT Press
- de Graaf MMA (2016) An ethical evaluation of human-robot relationships. *Int J Soc Robot* 8(4):589–598. <https://doi.org/10.1007/s12369-016-0368-5>
- Decker M (2008) Caregiving robots and ethical reflection: the perspective of interdisciplinary technology assessment. *AI Society* 22(3):315–330
- Duska R, Whelan M (1975) *Moral development: a guide to Piaget and Kohlberg*. Paulist Press
- Fischinger D, Einramhof P, Papoutsakis K, Wohlking W, Mayer P, Panek P, Hofmann S, Koertner T, Weiss A, Argyros A, Vincze M (2016) Hobbitt, a care robot supporting independent living at home: first prototype and lessons learned. *Robot Auton Syst* 75:60–78. <https://doi.org/10.1016/j.robot.2014.09.029>
- Fogg B, Nass C (1997) How users reciprocate to computers I CHI '97 extended abstracts on human factors in computing systems. *Extended Abst Hum Factors Comput Syst NY*. <https://doi.org/10.1145/1120212.1120419>
- Frennert S, Östlund B (2014) Review: seven matters of concern of social robots and older people. *Int J Soc Robot* 6(2):299–310. <https://doi.org/10.1007/s12369-013-0225-8>
- Gouldner AW (1960) The norm of reciprocity: a preliminary statement. *Am Sociol Rev* 25(2):161–178. <https://doi.org/10.2307/2092623>
- Kahn PH Jr, Friedman B, Perez-Granados DR, Freier NG (2004) Robotic pets in the lives of preschool children. *Extended Abstracts Hum Factors Comput Syst* 5:1449–1452. <https://doi.org/10.1145/985921.986087>
- Kahn, P. H., Ishiguro, H., Friedman, B., Kanda, T. (2006). What is a Human? - Toward psychological benchmarks in the field of human-robot interaction. *ROMAN 2006 - The 15th IEEE international symposium on robot and human interactive communication*, 364–371. <https://doi.org/10.1109/ROMAN.2006.314461>
- Lammer L, Huber A, Weiss A, Vincze M (2014) Mutual Care: How older adults react when they should help their care robot. *Proceedings of the 3rd International Symposium on New Frontiers in Human-Robot Interaction*, 1–4.
- Lorenz T, Weiss A, Hirche S (2016) Synchrony and reciprocity: key mechanisms for social companion robots in therapy and care. *Int J Soc Robot* 8(1):125–143. <https://doi.org/10.1007/s12369-015-0325-8>
- Oost E, Reed D (2011) Towards a sociological understanding of robots as companions. In: Lamers MH, Verbeek FJ (eds) *Human-robot personal relationships*. Springer
- Paiva A, Santos FP, Santos FC (2018) Engineering Pro-Sociality With Autonomous Agents. Thirty-Second aaai conference on artificial intelligence. thirty-second AAAI conference on artificial intelligence. <https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16799>
- Sander-Staudt M (2015) Caring reciprocity as a relational and political ideal in Confucianism and care ethics. In: Engelens D, Hamington M (eds) *Care ethics and political theory*. Oxford University Press
- Sandoval EB, Brandstetter J, Obaid M, Bartneck C (2016) Reciprocity in human-robot interaction: a quantitative approach through the prisoner's dilemma and the ultimatum game. *Int J Soc Robot* 8(2):303–317. <https://doi.org/10.1007/s12369-015-0323-x>
- Scheutz M (2012) The inherent dangers of unidirectional emotional bonds between humans and social robots. In: Lin P, Abney K, Bekey G (eds) *Robot ethics: the ethical and social implications of robotics*. MIT Press
- Sharkey A (2014) Robots and human dignity: a consideration of the effects of robot care on the dignity of older people. *Ethics Inf Technol* 16(1):63–75. <https://doi.org/10.1007/s10676-014-9338-5>
- Sharkey A, Sharkey N (2012) Granny and the robots: ethical issues in robot care for the elderly. *Ethics Inf Technol* 14(1):27–40
- Sullins JP (2012) Robots, love, and sex: the ethics of building a love machine. *IEEE Trans Affect Comput* 3(4):398–409. <https://doi.org/10.1109/T-AFFC.2012.31>
- Tronto J (1993) *Moral boundaries: a political argument for an ethic of care*. Routledge
- van Wynsberghe A (2013a) A method for integrating ethics into the design of robots. *Industrial Robot An Internat J* 40(5):433–440. <https://doi.org/10.1108/IR-12-2012-451>
- van Wynsberghe A (2013b) Designing robots for care: care centered value-sensitive design. *Sci Eng Ethics* 19(2):407–433. <https://doi.org/10.1007/s11948-011-9343-6>
- van Wynsberghe A (2016) Service robots, care ethics, and design. *Ethics Inf Technol* 18(4):311–321. <https://doi.org/10.1007/s10676-016-9409-x>
- van Wynsberghe A, Li S (2019) A paradigm shift for robot ethics: from HRI to human–robot–system interaction (HRSI). *Medicol Bioethics* 9:11–20. <https://doi.org/10.2147/MB.S160348>
- Vanlaere L, Gastmans C (2011) A personalist approach to care ethics. *Nurs Ethics Nursing Ethics* 18(2):161–173
- Verkerk M (2001) The care perspective and autonomy. *Med Health Care Philos* 4(3):289–294

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.