

Ethics and social robotics

Raffaele Rodogno¹

Published online: 18 November 2016
© Springer Science+Business Media Dordrecht 2016

On the Western traditional model of subjectivity, with its roots in Descartes and Kant, self-consciousness, rationality, individuality, freedom, agency, responsibility, and moral dignity come as a package deal. This package is presupposed by any relation or interaction that deserves the label ‘social’. Moreover, on the Hegelian reworking of this model, the package of capacities called subjectivity in turn presupposes social relations. On this model only humans are the kind of entity that can stand in social relations, and standing in social relations confers these human capacities and the rights and statuses that adhere to them.

While this model has been challenged from within by both the ‘analytical’ and the ‘continental’ philosophical traditions, social robotics challenges the traditional link between subjectivity and sociality from a new angle. It seems that robots have some of what it takes to be social agents, at least if we take at face value the way in which we (are willing to) interact with them, as if they were real pets, or confidants, or friends. What are we to make of that? Should we extend our concepts and adjust the definitional conditions for social agency? But how will this change propagate through the network of our foundational concepts? Or should we insist that talk about ‘social robotics’ is question-begging? But how to exclude robots from the community of thinkers, now, when we have accustomed ourselves to describe the human mind in terms evolutionary algorithms, neural nets, dynamic systems, complexity, Bayesian updating, mechanisms; have we not roboticized

the human mind to an extent that we are now forced to consider robots our functional equals, in principle at least?¹

These considerations and the questions to which they give rise form the background to the conference *Robo-Philosophy 2014—Sociable Robots and the Future of Social Relations* (August 20–23, Aarhus University, Denmark), of which this special issue is an offspring.^{2,3} The conference—and its theoretical background—was conceived by Johanna Seibt as the first of a series of five biannual conferences intended to find the right answers to the questions above. In practice, the conference functioned as a catalyst to the idea that philosophers should join the already existing efforts of roboticists in exploring artificial social agency. More in particular, the conference showed that philosophical reflection on social robotics pertains to all systematic areas of philosophy, not only to ethics, which had engaged this theme for some time, but also to philosophy of cognition, ontology, epistemology, philosophy of science, philosophical anthropology and aesthetics, philosophy of culture and intercultural philosophy, political philosophy, and even to philosophy of religion. In fact, since the 2014 conference, it has become clear that to understand the transformative potentials of human-robot interactions, the rapid development of social robotics calls for an integrated effort not just in philosophy but across the Humanities.⁴

✉ Raffaele Rodogno
filrr@cas.au.dk

¹ Department of Philosophy and History of Ideas, Aarhus University, Jens Chr. Skous Vej 7, Building 1467/417, 8000 Århus C, Denmark

¹ The wording of the first two paragraphs follows closely Seibt et al. (2014 p. viii).

² Other offspring are Nørskov (2015), Hakli and Seibt (2016), Seibt et al. (2017).

³ To be precise, five out of the nine articles collected here were originally presented at the conference.

⁴ Hence, the thoroughgoing interdisciplinary focus and discussions that characterize the second conference of the series, *Robophilosophy 2016—What Social Robots Can and Should Do* (October 17–21, Aarhus University, Denmark).

That this development *has to* happen rapidly is indeed impressed on us by some key facts. The robotics industry is currently growing rapidly—a recent study by the McKinsey Global Institute (2013) predicts that the market value for advanced robotics will be up to 17 trillion US \$ per year by 2025. Moreover, the development of this market sector is to a large extent driven by new applications in the area of ‘social robotics.’ For illustration from the European context, the research agenda “Robotics2020” of *euRobotics*, an association of 183 European robotics firms and research institutes, predicts that by 2020 robotics will “influence every aspect of work and home.” Since ‘social robots’ are designed to enter the space of human social interaction both physically and semantically, presenting a new type of ‘social’ agent, ‘social robotics’ has been aptly classified as “disruptive” technology, i.e., as a sort of technology that affects the core of our current social practices and might lead to profound cultural change.

In the light of these facts, it seems indeed important to have this volume focus on the ethical issues posed by social robots. At the most general level, these issues can be divided in two groups: (1) the scope or limits of morality, i.e., questions concerning the aptness of social robots as moral agents and/or as the objects of moral consideration; and (2) substantive normative questions, i.e., questions about the desirability or permissibility of designing, producing and/or deploying social robots in specific interaction contexts. Clearly, these two groups of issues are tightly connected: specific conceptualizations of morality and its limits will have specific ramifications on what is to count or not to count as a legitimate ethical question. While many of the following contributions do in fact mirror the connection between these two themes, at time moving from concerns pertaining to the one to those pertaining to the other, it is safe to say that each contribution focusses primarily on either one of them.

The contributions of Martin Bentzen, Katharyn Hogan, Migle Laukyte, Bertram Malle, and Niklas Toivakainen fall more squarely in the first group. Malle, for one, does interesting work on the idea of moral competence in robots, while discussing a number of ethical questions concerning the design, use, and treatment of moral robots in society. Laukyte argues that artificial non-human entities such as robots can be legitimate bearers of rights because they instantiate the same type of agency as other non-human entities such as group agents that are already considered as rights-bearers. Bentzen also focuses on the notion of agency and, in particular, takes issue with the claim that this capacity is instantiated by robotic systems consisting of a neuron culture grown on a multielectrode array (MEA). Finally, both Hogan and Toivakainen, while broadly sharing the phenomenological outlook appealed to

by authors such as Gunkel (2014) and Coeckelbergh (2014), reject (for different reasons) the conclusions arrived at by these authors concerning the moral status of robots.

The contributions of John Danaher, Raffaele Rodogno, Amanda Sharkey, and Aimee van Wynsberghe fall more squarely in the second group. The interaction contexts on which each contribution focusses and the type of substantive ethical issue at play do however vary from one contribution to the other. Danaher’s article, for one, does not focus on any specific context of human-robot interaction but is rather interested in all those interactions, whichever their context, in which questions of (robot-directed) blame may arise. Danaher argues that with the increasing introduction of robots in society a gap may arise between the human desire for retribution and the absence of appropriate subjects of retributive blame. This gap, he argues, has negative social and moral implications. In Sharkey’s article, classrooms are the context at issue, and privacy, attachment, deception, loss of human contact, control, and accountability the relevant ethical issues. Sharkey argues that there are good reasons not to welcome fully fledged robot teachers and that robot companions in the classroom should be given a cautious welcome at best. Wynberghe focuses on personal and professional service robots and puts forward a care ethics approach aimed at incorporating ethics into the design process of robots. Finally, Rodogno examines the nature of human-robot pet relations that appear to involve genuine affective responses on behalf of humans and argues that such relations do not necessarily involve (self-)deception and, even if they did, that the sentimentality involved by them would not be morally objectionable.

References

- Coeckelbergh, M. (2014). The moral standing of machines: Towards a relational and non-Cartesian moral hermeneutics. *Philosophy & Technology*, 27(1), 61–77.
- Gunkel, D. J. (2014). A vindication of the rights of machines. *Philosophy & Technology*, 27(1), 113–132.
- Hakli, R., & Seibt, J. (Eds.). (2016). *Sociality and normativity for robots—philosophical investigations*. New York: Springer. **forthcoming**.
- McKinsey Global Institute (2013) Disruptive technologies: Advances that will transform life, business, and the global economy. http://www.mckinsey.com/insights/business_technology/disruptive_technologies.
- Nørskov, M. (Ed.). (2015). *Sociable robots—boundaries, potentials, challenges*. Farnham: Ashgate.
- Seibt, J., Hakli, R., & Nørskov, M. (Eds.). (2014). *Sociable robots and the future of social relations*. Amsterdam: IOS Press.
- Seibt, J., Hakli, R., & Nørskov, M. (Eds.). (2017). *Robophilosophy—philosophy of, for, and by social robotics*. Cambridge: MIT Press.