Laukyte, M. (2017). Social Robots: Boundaries, Potential, Challenges [book review]. *NanoEthics*, 11 (3), pp. 273-275.

DOI: 10.1007/s11569-017-0291-8

Social Robots: Boundaries, Potential, Challenges

Marco Nørskov (ed.) 2016 (Dorchester, Ashgate) ISBN-13: 978-1472474308. 244pp.

Migle Laukyte

Departamento de Derecho Privado, Universidad Carlos III de Madrid, calle Madrid 126, 28903 Getafe, Madrid, Spain
e-mail: mlaukyte@der-pr.uc3m.es

Mark Twain once said that a classic is a book which people praise and don't read. If this is true, then I do not want this book to be a classic. I want it to be read and I will use the rest of this book review to convince you to do so.

But before I start, a premise should be made: I assume that if you are reading this book review, then you are a philosopher, a sociologist, a legal scholar, an anthropologist, a culture studies scholar, a communications expert, or anyone with an interest in the future of technology, and in particular artificial intelligence (AI), robotics, and the different forms that human-AI or robot interaction can take. Among the possible readers of this book, I would also include people who actually design and build intelligent machines, and so if you are in one of these groups or consider yourself to be, then this book is for you.

It was written by people like you who in the summer of 2014 met for a Robo-philosophy conference in the windy city of Aarhus (Denmark) to discuss how we could or should frame our relationships with a particular kind of robots, namely, social robots—robots that are built to closely interact with us, our children, our parents, friends, patients, clients, and pets. This book is an outcome of the conference.

Of course, this book is not and cannot be exhaustive in treating the issue, and the editor cautions us in that regard: Here, I think, lies the difficulty of Marco Nørskov's task—the difficulty of selecting the works that would be included in this volume. The 2014 Robo-philosophy conference included more than one hundred paper presentations, and it goes without saying that selecting twelve articles from such a large pool is a challenging enterprise, one that Nørskov accomplished with success.

The book, as its title suggests, is organized into three parts. The first part, *Boundaries*, explores the limits or boundaries that for the time being differentiate human–social robot interaction from human–human interaction. Julia Knifa ("On the significance of understanding in human robot interaction") argues that our human social interaction is based on self-understanding, and this is not yet possible in our interaction with social robots, since these robots cannot be said to properly understand us, nor can they be said to have a notion of themselves or of their interaction with us. The same applies to the emotions and to humanly recognizable interests, which according to Raffaele Rodogno ("Robots and the limits of morality") are the *condicio sine qua non* for becoming subject to moral consideration, a condition that robots do not yet possess.

Refreshingly surprising in this section dedicated to boundaries is Josh Redstone's discussion of the complex nature of the feeling of empathy ("Making sense of empathy with sociable robots"), where instead of reflections on the inability of robots to feel empathy, we find a story about our failure to feel genuine empathy toward machines. I find this a very interesting idea, fleshed out by bringing in a number of elements, including the uncanny valley phenomenon. So the boundaries discussed in this part of the book are not only those that

social robots have to overcome so as to become true participants in social interaction, but also those that we human beings have, and it is this dual perspective that enables this book to bring fresh insights to the question of the human–robot interaction.

The second part, *Potential*, is devoted to the potential of social robotics. This potential can be hindered by an almost instinctive reluctance of engineers to subject their research to ethical review, and John P. Sullins ("Ethical boards for research in robotics and artificial intelligence: is it too soon to act?") makes the case that this reluctance will not help the effort to improve the design and functionality of social robots. Robotics and artificial intelligence have a serious image problem in the media, and this is why ethics could be of use here in solving this problem. Due to the close interaction between robots and human beings, firms can no longer think only in the categories such as "profitable", "cost effective", and "safe or unsafe" but should also think in terms of "right and wrong", and this is where ethical boards come into play. The discussion here is particularly useful because it provides some practical ideas on how an ethical board could foster a dialogue with industry.

Also worthy of mention in this regard is Gunhild Borggreen's discussion ("Staging lies: performativity in the human-robot theatre play *I, Worker*") of the robot-human theatre play *Hataraku Watashi (I, Worker)*. It provides us with another example of the potential of social robots, namely that they can become works of art in their own right, telling us more about ourselves and thus helping software engineers and other scientists to use this knowledge so as to rethink and improve the existing design for human–robot interaction. In this sense, the theatre becomes a robotics lab where the experiment of interacting with robots is not only performed on stage but also brought to the public.

The second part of the book, *Potential*, is linked to the first part, *Boundaries*, by means of the attention devoted to the uncanny valley phenomenon. But while in *Boundaries* the uncanny valley serves to illustrate how we fail to empathize with social robots, in *Potential*, it becomes a key to revealing to us something about human nature, and in particular about the process of dehumanization, which is what leads us to experience the uncanny valley.

The third part of the book, *Challenges*, addresses four challenges that social robots give rise to:

1. The challenge posed to the modern concept of social actors (and of sociality itself). Social robots have a deep impact on the institutional order of human society, affecting our perception of what a legitimate social actor is and how this sphere ought to be delimited. Social robots challenge the modern concept of personhood—the notion that only human beings are persons—thereby also calling into question the distinction between humans and machines, due to their capacity for interaction, autonomy, and learning.
2. The challenge posed to the social ontology through which we assess who or what can be held morally and legally responsible. This challenge can be constructed as a specification of the first one, but it also adds a complicating condition, for the problem concerns not just social robots but also "nonlocalizable" ones, that is, robots made of millions of dynamic components that are continuously replaced in the robot's body.
3. The challenge of gender ascription and its implications in social robotics. The problem here is that the way we interact with social robots (utilitarian or affective ones) will to some extent depend on how they look like, a factor that in subtle but significant ways may influence our expectations, in just the same way as it happens with humans, where we sometimes expect different people to behave differently simply because of their looks. Here, anthropomorphism comes into play—our tendency to anthropomorphize a robot helps us interact with it. The problem lies in the lack of objectivity both in the way we design social robots and the way we interact with them: Both their design and our interaction with them are shaped by our gender stereotypes, and these stereotypes may be reinforced by the interaction.
4. The challenge posed to our freedom of choice and action when we become objects of control. The problem here stems from the ability of robots to frame our interaction with them, in a situation where we either do what the smart tool or robot asks us to do or we stop interacting with it altogether. So what happens is that smart tools may "undermine one's personal sovereignty," and they can be especially successful in doing so if we are talking about smart tools (including social robots) designed for everyday life. This process of disempowerment has further implications as concerns our personal responsibility and accountability: To what extent can I still be held accountable for

my actions if my freedom to choose and act is constrained by my interaction with a robot?

The more we drill down into these issues, the more we realize that the distinction between the three parts into which the book is divided needs to be taken with caution: The line between boundaries, potential, and challenges is fuzzy and sometimes even difficult to see. The discussion on the role of understanding in human and robot interaction (Knifka's contribution in *Boundaries*), for example, overlaps with the discussion of objectified interaction frames (Hironori Matsuzaki's contribution in *Challenges*). The same can be said about the discussion of ethical boards (Sullins' contribution in *Potential*) and Matsuzaki's chapter

If you have made it to here, you may conclude that this book review is essentially a song of praise, and hence that the book may become a classic that no one will read. I do not think so. I think it *will* be read, because it presents social robotics as a lens through which to discover ourselves in a new light. In fact, as the editor insightfully remarks, unlike any other technology invented so far, social robots enable us to truly reflect on who we are. And this, I submit, is where the value of this book ultimately lies; it gives us a fresh perspective on social robots, prompting us to reflect on what it means to be human.