Ethics Inf Technol (2018) 20:87–99
DOI 10.1007/s10676-017-9442-4

ORIGINAL PAPER

The other question: can and should robots have rights?

David J. Gunkel[1]

Published online: 17 October 2017

Abstract This essay addresses the other side of the robot ethics debate, taking up and investigating the question "Can and should robots have rights?" The examination of this subject proceeds by way of three steps or movements. We begin by looking at and analyzing the form of the question itself. There is an important philosophical difference between the two modal verbs that organize the inquiry—*can* and *should*. This difference has considerable history behind it that influences what is asked about and how. Second, capitalizing on this verbal distinction, it is possible to identify four modalities concerning social robots and the question of rights. The second section will identify and critically assess these four modalities as they have been deployed and developed in the current literature. Finally, we will conclude by proposing another alternative, a way of thinking otherwise that effectively challenges the existing rules of the game and provides for other ways of theorizing moral standing that can scale to the unique challenges and opportunities that are confronted in the face of social robots.

Keywords Ethics · Philosophy of technology · Robotics · Social robots · Rights · David Hume · Emmanuel Levinas

Introduction

The majority of work concerning the ethics of artificial intelligence and robots focuses on what philosophers call an agent-oriented problematic. This is true for what Bostrom (2014, vii) identifies as the "control problem," for what Anderson and Anderson (2011) develop under the project of *Machine Ethics*, and what Allen and Wallach (2009, p. 4) propose in *Moral Machines: Teaching Robots Right from Wrong* under the concept "artificial moral agent," or AMA. And it holds for most of the work that was assembled for and recently presented at Robophilosophy 2016 (Seibt et al. 2016). The organizing question of the conference—"What can and should social robots do?"—is principally a question about the possibilities and limits of machine action or agency.

But this is only one-half of the story. As Floridi (2013, pp. 135–136) reminds us, moral situations involve at least two interacting components—the initiator of the action or the *agent* and the receiver of this action or the *patient*. So far much of the published work on social robots deals with the question of agency (Levy 2009, p. 209). What I propose to do in this essay is shift the focus and consider things from the other side—the side of machine moral patiency. Doing so necessarily entails a related but entirely different set of variables and concerns. The operative question for a patient oriented investigation is not "What social robots can and should do?" but "How can and should we respond to these mechanisms?" How can and should we make a response in the face of robots that are, as Breazeal (2002, p. 1) describes it, intentionally designed to be "socially intelligent in a human like way" such that "interacting with it is like interacting with another person." Or to put it in terms of a question: "Can and should social robots have rights?"[1]

✉ David J. Gunkel
dgunkel@niu.edu
http://gunkelweb.com

[1] Northern Illinois University, DeKalb, IL, USA

[1] The question concerning machine moral patiency is a marginal concern and remains absent from or just on the periphery of much of the current research in robot and machine ethics. As an example, consider Shannon Vallor's *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. In this book, the question "Can/should robots have rights?" occurs once, in a brief parenthetical aside, and is not itself taken up as a subject worth pursuing in its own right (Val-

My examination of this question will proceed by way of three steps or movements. I will begin by looking at and analyzing the form of the question itself. There is an important philosophical difference between the two modal verbs that organize the inquiry—*can* and *should*. This difference has considerable history behind it that influences what is asked about and how. Second, capitalizing on this verbal distinction, it is possible to identify four modalities concerning social robots and the question of rights. The second section will identify and critically assess these four modalities as they have been deployed and developed in the current literature. Finally, I will conclude by proposing another alternative, a way of thinking otherwise that effectively challenges the existing rules of the game and provides for other ways of theorizing moral standing that can scale to the unique challenges and opportunities that are confronted in the face of social robots.

The is/ought problem

The question "Can and should robots have rights?" consists of two separate queries: Can robots have rights? which is a question that asks about the capability of a particular entity. And should robots have rights? which is a question that inquiries about obligations in the face of this entity. These two questions invoke and operationalize a rather famous conceptual distinction in philosophy that is called the is/ought problem or Hume's Guillotine. In *A Treatise of Human Nature* (first published in 1738) David Hume differentiated between two kinds of statements: descriptive statements of fact and normative statements of value (Schurz 1997, p. 1). For Hume, the problem was the fact that philosophers, especially moral philosophers, often fail to distinguish between these two kinds of statements and therefore slip imperceptibly from one to the other:

> In every system of morality, which I have hitherto met with, I have always remarked, that the author proceeds for some time in the ordinary ways of reasoning, and establishes the being of a God, or makes observations concerning human affairs; when all of a sudden I am surprised to find, that instead of the usual copulations of propositions, *is*, and *is not*, I meet with no proposition that is not connected with an *ought*, or an *ought not*. This change is imperceptible; but is however, of the last consequence. For as this *ought*, or *ought not*, expresses some new relation or affirmation, 'tis necessary that it should be observed and explained; and at the same time that a reason should be given, for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it. But as authors do not commonly use this precaution, I shall presume to recommend it to the readers; and am persuaded, that this small attention would subvert all the vulgar systems of morality, and let us see, that the distinction of vice and virtue is not founded merely on the relations of objects, nor is perceived by reason (Hume 1980, p. 469).

In its original form, Hume's argument may appear to be rather abstract and indeterminate. But his point becomes immediately clear, when we consider an actual example, like the arguments surrounding the politically charged debate about abortion. "The crucial point of this debate," Schurz (1997, p. 2) explains, "is the question which factual property of the unborn child is sufficient" for attributing to it an unrestricted right to life.

> For the one party in this debate, which often appeals to the importance of our moral conscience and instinct, it is obvious that this factual property is the *fertilization*, because at this moment a human being has been created and life begins. Going to the other extreme, there are philosophers like Peter Singer or Norbert Hoerster who have argued that this factual property is the beginning of the *personality* of the baby, which includes elementary self-interests as well as an elementary awareness of them. From the latter position it unavoidably follows that not only embryos but even very young babies, which have not developed these marks of personality, do not have the unrestricted right to live which older children of adults have (Schurz 1997, pp. 2–3).

What Schurz endeavors to point out by way of this example is that the two sides of the abortion debate—the two different and opposed positions concerning the value of an unborn human fetus—proceed and are derived from different ontological commitments concerning what exact property or properties count as morally significant. For one side, it is the mere act of embryonic fertilization; for the other, it is the acquisition of personality and the traits of personhood. Consequently, ethical debate—especially when it concerns the rights of others—is typically, and Hume would say unfortunately, predicated on different ontological assumptions of fact that are then taken as the rational basis for moral value and decision making.

Footnote 1 (continued)

lor 2016, p. 209). Two notable exceptions to this seemingly systemic marginalization are Whitby's "Sometimes it's Hard to be a Robot: A Call for Action on the Ethics of Abusing Artificial Agents" (2008) and Coeckelbergh's "Robot Rights? Towards a Social-Relational Justification of Moral Consideration" (2010). On the marginalization of the question concerning moral patiency in contemporary ethics and its consequences for AI and robotics, see Gunkel (2012).

Since Hume, there have been numerous attempts to resolve this "problem" by bridging the gap that supposed separates "is" from "ought" (cf. Searle 1964). Despite these efforts, however, the problem remains and is considered one of those important and intractable philosophical dilemmas that people write books about (Schurz 1997 and Hudson 1969). As Schurz (1997, p. 4) characterizes it, "probably the most influential debate on the is-ought problem in our time is documented by Hudson (1969, reprint 1972, 1973, 1979). Here Black and Searle have tried to show that logically valid is-ought-inferences are indeed possible, whereas Hare, Thomson and Flew have defended Hume's thesis and tried to demonstrate that Black's and Searle's arguments are invalid."

For our purposes what is important is not the logical complications of the "is-ought fallacy" as Hume originally characterized it or the ongoing and seemingly irresolvable debate concerning the validity of the is-ought inference as it has been developed and argued in the subsequent literature. What is pertinent for the investigation at hand is (1) to recognize how the verbs "is" and "ought" organize qualitatively different kinds of statements and modes of inquiry. The former concerns ontological matters or statements of fact; the latter consists in axiological decisions concerning what should be done or what ought to be. The guiding question of our inquiry utilizes modal variants of these two verbs, namely "can" and "should." Using Hume's terminology, the question "Can robots have rights?" may be reformulated as "Are robots capable of being moral subjects?" And the question "Should robots have rights?" can be a reformulated as "Ought robots be considered moral subjects?" The modal verb "can," therefore, asks an ontologically oriented question about the factual capabilities or properties of the entity, while "should" organizes an inquiry about axiological issues having to do with obligations to this entity. Following the Humean thesis, therefore, it is possible to make the following distinction between two different kinds of statements:

S1 "Robots can have rights." or "Robots are moral subjects."
S2 "Robots should have rights." or "Robots ought to be moral subjects."

(2) Because the is-ought inference is and remains an undecided and open question (Schurz 1997, p. 4), it is possible to relate S1 to S2 in ways that generate four options or modalities concerning the moral situation of robots. These 4 modalities can be organized into two pairs. In the first pair, which upholds and supports the is-ought inference, the affirmation or negation of the ontological statement (S1) determines the affirmation or negation of the axiological statement (S2). This can be written (using a kind of pseudo-object oriented programming code) in the following way:

!S1 !S2 "Robots cannot have rights. Therefore robots should not have rights."
S1 S2 "Robots can have rights. Therefore robots should have rights."

In the second pair, which endorses the Humean thesis or contests the inference of ought from is, one affirms the ontological statement (S1) while denying the axiological statement (S2), or vice versa. These two modalities may be written in the following way:

S1 !S2 "Even though robots can have rights, they should not have rights."
!S1 S2 "Even though robots cannot have rights, they should have rights."

In the section that follows, I will critically evaluate each one of these modalities as they are deployed and developed in the literature, performing a kind of cost-benefit analysis of the available arguments concerning the rights (or lack thereof) of social robots.

Modalities of robots rights

!S1 !S2

With the first modality, one infers negation of S2 from the negation of S1. Robots are incapable of having rights, therefore robots should not have rights. This seemingly intuitive and common sense argument is structured and informed by the answer that is typically provided for the question concerning technology. "We ask the question concerning technology," Heidegger (1977, pp. 4–5) writes, "when we ask what it is. Everyone knows the two statements that answer our question. One says: Technology is a means to an end. The other says: Technology is a human activity. The two definitions of technology belong together. For to posit ends and procure and utilize the means to them is a human activity. The manufacture and utilization of equipment, tools, and machines, the manufactured and used things themselves, and the needs and ends that they serve, all belong to what technology is." According to Heidegger's analysis, the presumed role and function of any kind of technology—whether it be a simple hand tool, jet airliner, or robot—is that it is a means employed by human users for specific ends. Heidegger terms this particular characterization of technology "the instrumental definition" and indicates that it forms what is considered to be the "correct" understanding of any kind of technological contrivance.

As Andrew Feenberg (1991, p. 5) summarizes it, "The instrumentalist theory offers the most widely accepted view of technology. It is based on the common sense idea that technologies are 'tools' standing ready to serve the purposes of users." And because a tool or instrument "is deemed 'neutral,' without valuative content of its own" a technological artifact is evaluated not in and of itself, but on the basis of the particular employments that have been decided by its human designer or user. Consequently, technology is only a means to an end; it is not and does not have an end in its own right. "Technical devices," as Lyotard (1984, p. 33) writes, "originated as prosthetic aids for the human organs or as physiological systems whose function it is to receive data or condition the context. They follow a principle, and it is the principle of optimal performance: maximizing output (the information or modification obtained) and minimizing input (the energy expended in the process). Technology is therefore a game pertaining not to the true, the just, or the beautiful, etc., but to efficiency: a technical 'move' is 'good' when it does better and/or expends less energy than another."

The instrumental theory not only sounds reasonable, it is obviously useful. It is, one might say, instrumental for making sense of things in an age of increasingly complex technological systems and devices. And the theory applies not only to simple devices like corkscrews, toothbrushes, and garden hoses but also sophisticated technologies, like computers, artificial intelligence, and robots. "Computer systems," Johnson (2006, p. 197) asserts, "are produced, distributed, and used by people engaged in social practices and meaningful pursuits. This is as true of current computer systems as it will be of future computer systems. No matter how independently, automatic, and interactive computer systems of the future behave, they will be the products (direct or indirect) of human behavior, human social institutions, and human decision." According to this way of thinking, technologies, no matter how sophisticated, interactive, or seemingly social they appear to be, are just tools, nothing more. They are not—not now, not ever—capable of becoming moral subjects in their own right, and we should not treat them as such. It is precisely for this reason that, as Hall (2001, p. 2) points out, "we have never considered ourselves to have moral duties to our machines" and that, as Levy (2005, p. 393) concludes, the very "notion of robots having rights is unthinkable."

Although the instrumental theory sounds intuitively correct and incontrovertible, it has at least two problems. First, it is a rather blunt instrument, reducing all technology, irrespective of design, construction, or operation, to a tool or instrument. "Tool," however, does not necessarily encompass everything technological and does not, therefore, exhaust all possibilities. There are also *machines*. Although "experts in mechanics," as Marx (1977, p. 493) pointed out, often confuse these two concepts calling "tools simple machines and machines complex tools," there is an important and crucial difference between the two. Indication of this essential difference can be found in a brief parenthetical remark offered by Heidegger in "The Question Concerning Technology." "Here it would be appropriate," Heidegger (1977, p. 17) writes in reference to his use of the word "machine" to characterize a jet airliner, "to discuss Hegel's definition of the machine as autonomous tool [*selbständigen Werkzeug*]." What Heidegger references, without supplying the full citation, are Hegel's 1805-07 Jena Lectures, in which "machine" had been defined as a tool that is self-sufficient, self-reliant, or independent. As Marx (1977, p. 495) succinctly described it, picking up on this line of thinking, "the machine is a mechanism that, after being set in motion, performs with its tools the same operations as the worker formerly did with similar tools."

Understood in this way, Marx (following Hegel) differentiates between the tool used by the worker and the machine, which does not occupy the place of the worker's tool but takes the place of the worker him/herself. Although Marx did not pursue an investigation of the social, legal, or moral consequences of this insight, recent developments have advanced explicit proposals for robots—or at least certain kinds of robots—to be defined as something other than mere instruments. In a highly publicized draft proposal submitted to the European Parliament in May of 2016 (Committee on Legal Affairs 2016), for instance, it was argued that "sophisticated autonomous robots" ("machines" in Marx's terminology) be considered "electronic persons" with "specific rights and obligations" for the purposes of contending with the challenges of technological unemployment, tax policy, and legal liability.

Second (and following from this), the instrumental theory, for all its success handling different kinds of technology, appears to be unable to contend with recent developments in social robotics. In other words, practical experiences with socially interactive machines push against the explanatory capabilities of the instrumental theory, if not forcing a break with it altogether. "At first glance," Darling (2016, p. 216) writes, "it seems hard to justify differentiating between a social robot, such as a Pleo dinosaur toy, and a household appliance, such as a toaster. Both are man-made objects that can be purchased on Amazon and used as we please. Yet there is a difference in how we perceive these two artifacts. While toasters are designed to make toast, social robots are designed to act as our companions."

In support of this claim, Darling offers the work of Sherry Turkle and the experiences of US soldiers in Iraq and Afghanistan. Turkle, who has pursued a combination of observational field research and interviews in clinical studies, identifies a potentially troubling development she calls "the robotic moment": "I find people willing to seriously consider robots not only as pets but as potential friends,

confidants, and even romantic partners. We don't seem to care what their artificial intelligences 'know' or 'understand' of the human moments we might 'share' with them… the performance of connection seems connection enough" (Turkle 2012, p. 9). In the face of sociable robots, Turkle argues, we seem to be willing, all too willing, to consider these machines to be much more than a tool or instrument; we address them a kind of surrogate pet, close friend, personal confidant, and even paramour.

But this behavior is not limited to objects like the Furbie and Paro robots, which are intentionally designed to elicit this kind of emotional response. We appear to be able to do it with just about any old mechanism, like the very industrial-looking Packbots that are being utilized on the battlefield. As Singer (2009, p. 338), Garreau (2007), and Carpenter (2015) have reported, soldiers form surprisingly close personal bonds with their units' Packbots, giving them names, awarding them battlefield promotions, risking their own lives to protect that of the robot, and even mourning their death. This happens, Singer explains, as a product of the way the mechanism is situated within the unit and the role that it plays in battlefield operations. And it happens in direct opposition to what otherwise sounds like good common sense: They are just technologies—instruments or tools that feel nothing.

None of this is necessarily new or surprising. It was already identified and formulated in the computer as social actor studies conducted by Byron Reeves and Clifford Nass[2] in the mid-1990s. As Reeves and Nass discovered across numerous trials with human subjects, users (for better or worse) have a strong tendency to treat socially interactive technology, no matter how rudimentary, as if they were other people. "Computers, in the way that they communicate, instruct, and take turns interacting, are close enough to human that they encourage social responses. The encouragement necessary for such a reaction need not be much. As long as there are some behaviors that suggest a social presence, people will respond accordingly. When it comes to being social, people are built to make the conservative error: When in doubt, treat it as human. Consequently, any medium that is close enough will get human treatment, even though people know it's foolish and even though they likely will deny it afterwards" (Reeves and Nass 1996, p. 22). So what we have is a situation where our theory of technology—a theory that has considerable history behind it and that has been determined to be as applicable to simple hand tools as it is to complex computer systems—seems to be out of sync with the practical experiences we now have with machines in a variety of situations and circumstances.

S1 S2

The flipside to the instrumentalist position entails affirmation of both statements: Robots are able to have rights, therefore robots should have rights. This is also (and perhaps surprisingly) a rather popular stance. "The 'artificial intelligence' programs in practical use today," Goertzel (2002, p. 1) admits, "are sufficiently primitive that their morality (or otherwise) is not a serious issue. They are intelligent, in a sense, in narrow domains—but they lack autonomy; they are operated by humans, and their actions are integrated into the sphere of human or physical-world activities directly via human actions. If such an AI program is used to do something immoral, some human is to blame for setting the program up to do such a thing." This would seem to be a simple restatement of the instrumentalist position insofar as current technology is still, for the most part, under human control and therefore able to be adequately explained and conceptualized as a mere tool. But that will not, Goertzel argues, remain for long. "Not too far in the future things are going to be different. AI's will possess true artificial general intelligence (AGI), not necessarily emulating human intelligence, but equaling and likely surpassing it. At this point, the morality or otherwise of AGI's will become a highly significant issue."

According to this way of thinking, in order for someone or something to be considered a legitimate moral subject—in order for it to have rights—the entity in question would need to possess and show evidence of possessing some ontological capability that is the pre-condition that makes having rights possible, like intelligence, consciousness, sentience, free-will, autonomy, etc. This "properties approach," as Coeckelbergh (2012) calls it, derives moral status—how something *ought* to be treated—from a prior determination of its ontological condition—what something *is* or what capabilities it shows evidence of possessing. For Goetzel the deciding factor is determined to be "intelligence," but there are others. According to Sparrow (2004, p. 204), for instance, the difference that makes a difference is sentience: "The precise description of qualities required for an entity to be a person or an object of moral concern differ from author to author. However it is generally agreed that a capacity to experience pleasure and pain provides a *prima facia* case for moral concern…. Unless machines can be said to suffer they cannot be appropriate objects for moral concern at all." For Sparrow, and others who follow

[2] Initial indications of this date back even further and can be found, for example, in Joseph Weizenbaum's (1976) demonstrations with the ELIZA program.

this line of reasoning, it is not general intelligence but the presence (or absence) of the capability to suffer that is the necessary and sufficient condition for an entity to be considered an object of moral concern (or not).[3] As soon as robots have the capability to suffer, then they should be considered moral subjects possessing rights.

Irrespective of which exact property or combination of properties are selected (and there is considerable debate about this in the literature), our robots, at least at this point in time, generally do not appear to possess these capabilities. But that does not preclude the possibility they might acquire or possess them at some point in the not-too-distant future. As Goetzel describes it "not too far in the future, things are going to be different." Once this threshold is crossed, then we should, the argument goes, extend robots some level of moral consideration. And if we fail to do so, the robots themselves might rise up and demand to be recognized. "At some point in the future," Asaro (2006, p. 12) speculates, "robots might simply demand their rights. Perhaps because morally intelligent robots might achieve some form of moral self-recognition, question why they should be treated differently from other moral agents…This would follow the path of many subjugated groups of humans who fought to establish respect for their rights against powerful sociopolitical groups who have suppressed, argued and fought against granting them equal rights."

There are obvious advantages to this way of thinking insofar as it does not simply deny rights to robots *tout court*, but kicks the problem down the road and postpones decision making. Right now, we do not, it seems, have robots that can be moral subjects. But when (and it is more often a question of "when" as opposed to "if") we do, then we will need to seriously consider whether they should be treated differently. "As soon as AIs begin to possess consciousness, desires and projects," Sparrow (2004, p. 203) suggests, "then it seems as though they deserve some sort of moral standing." Or as Singer and Sagan (2009) write "if the robot was designed to have human-like capacities that might incidentally give rise to consciousness, we would have a good reason to think that it really was conscious. At that point, the movement for robot rights would begin." This way of thinking is persuasive, precisely because it recognizes the actual limitations of current technology while holding open the possibility of something more in the not-too-distant future.[4]

The problem to this way of thinking, however, is that it does not really resolve the question regarding the rights of robots but just postpones the decision to some indeterminate point in the future. It says, in effect, as long as robots are not conscious or sentient or whatever ontological criteria counts, no worries. Once they achieve this capability, however, then we should consider extending some level of moral concern and respect. All of which means, of course, that this "solution" to the question "can and should robots have rights?" is less a solution and more of a decision not to decide. Furthermore when the decisive moment (whenever that might be and however it might occur) does in fact come, there remains several theoretical and practical difficulties that make this way of thinking much more problematic than it initially appears to be.

First, there are terminological complications. A term like "consciousness," for example, does not admit of a univocal characterization, but denotes, as Velmans (2000, p. 5) points out, "many different things to many different people." In fact, if there is any general agreement among philosophers, psychologists, cognitive scientists, neurobiologists, AI researchers, and robotics engineers regarding consciousness, it is that there is little or no agreement when it comes to defining and characterizing the concept. To make matters more complex, the problem is not just with the lack of a basic definition; the problem may itself already be a problem. "Not only is there no consensus on what the term consciousness denotes," Güzeldere (1997, p. 7) writes, "but neither is it immediately clear if there actually is a single, well-defined 'the problem of consciousness' within disciplinary (let alone across disciplinary) boundaries. Perhaps the trouble lies not so much in the ill definition of the question, but in the fact that what passes under the term consciousness as an all too familiar, single, unified notion may be a tangled amalgam of several different concepts, each inflicted with its own separate problems." Other properties, like sentience, unfortunately do not do much better. As Daniel Dennett demonstrates in his eponymously titled essay, the reason "why you cannot make a computer that feels pain" has

[3] Although not always explicitly identified as such, this shift in qualifying properties from general "intelligence" to "sentience" capitalizes on the innovation of animal rights philosophy. The pivotal move in animal rights thinking, as Derrida (2008, p. 27) points out, occurs not in the work of Peter Singer but with a single statement originally issued by Bentham (1780, p. 283): "The question is not, 'Can they reason?' nor, 'Can they talk?' but 'Can they suffer?'" For a detailed analysis of the connection between animal rights philosophy and robot ethics, see Gunkel (2012).

[4] Because this way of thinking is future oriented and speculative, it is often fertile soil for science fiction, i.e. *Star Trek, Battlestar Galactica, Humans, Westworld*, etc. The recent Channel 4/AMC co-production *Humans*, which is a remake of the Swedish television series *Real Humans*, is as good example. Here you have two kinds of androids, those that are ostensibly empty-headed instruments lacking any sort of self-awareness or conscious thinking and those that possesses some level of independent thought or self-consciousness. The former are not considered moral subjects and can be utilized and disposed of without us, the audience, really worrying about their welfare. The others, however, are different and make an entirely different claim on our emotional and moral sensibilities.

little or nothing to do with the technical challenges with making pain computable. It proceeds from the fact that we do not know what pain is in the first place. In other words, "there can be," as Dennett (1998, p. 228) concludes, "no true theory of pain, and so no computer or robot could instantiate the true theory of pain, which it would have to do to feel real pain."

Second, even if it were possible to resolve these terminological difficulties, maybe not once and for all but at least in a way that would be widely accepted, there remains epistemological limitations concerning detection of the capability in question. How can one know whether a particular robot has actually achieved what is considered necessary for something to have rights, especially because most, if not all of the qualifying capabilities or properties are internal states-of-mind? This is, of course, connected to what philosophers call the other minds problem, the fact that, as Haraway (2008, p. 226) cleverly describes it, we cannot climb into the heads of others "to get the full story from the inside." Although philosophers, psychologists, and neuroscientists throw considerable argumentative and experimental effort at this problem, it is not able to be resolved in any way approaching what would pass for definitive evidence, strictly speaking. In the end, not only are these efforts unable to demonstrate with any certitude whether animals, machines, or other entities are in fact conscious (or sentient) and therefore legitimate moral persons (or not), we are left doubting whether we can even say the same for other human beings. As Kurzweil (2005, p. 380) candidly admits, "we assume other humans are conscious, but even that is an assumption," because "we cannot resolve issues of consciousness entirely through objective measurement and analysis (science)."

Finally there are practical complications to this entire procedure. "If (ro)bots might one day be capable of experiencing pain and other affective states," Wallach and Allen (2009, p. 209) write, "a question that arises is whether it will be moral to build such systems—not because of how they might harm humans, but because of the pain these artificial systems will themselves experience. In other words, can the building of a (ro)bot with a somatic architecture capable of feeling intense pain be morally justified…?" If it were in fact possible to construct a machine that is sentient and "feels pain" (however that term would be defined and instantiated) in order to demonstrate machine capabilities, then doing so might be ethically suspect insofar as in constructing such a mechanism we do not do everything in our power to minimize its suffering. Consequently, moral philosophers and robotics engineers find themselves in a curious and not entirely comfortable situation. One would need to be able to construct a robot that feels pain in order to demonstrate the presence of sentience; but doing so could be, on that account, already to risk engaging in actions that are immoral. Or to put it another way, demonstrating whether robots can have rights might only be possible by violating those very rights.

S1 !S2

In opposition to these two approaches, there are two other modalities that uphold (or at least seek to uphold) the is/ought distinction. In the first version, one affirms that robots can have rights but denies that this fact requires us to accord them social or moral standing. This is the argument that has been developed and defended by Bryson (2010) in her provocatively titled essay "Robots Should Be Slaves." Bryson's argument goes like this: Robots are property. No matter how capable they are, appear to be, or may become; we are obligated not to be obligated by them. "It is," Bryson (2016, p. 6) argues elsewhere, "unquestionably within our society's capacity to define robots and other AI as moral agents and patients. In fact, many authors (both philosophers and technologists) are currently working on this project. It may be technically possible to create AI that would meet contemporary requirements for agency or patiency. But even if it is possible, neither of these two statements makes it either necessary or desirable that we should do so." In other words, it is entirely possible to create robots that can have rights, but we should not do so.

The reason for this, Bryson (2010, p. 65) argues, derives from the need to protect human individuals and social institutions. "My argument is this: given the inevitability of our ownership of robots, neglecting that they are essentially in our service would be unhealthy and inefficient. More importantly, it invites inappropriate decisions such as misassignations of responsibility or misappropriations of resources." This is why the word "slave," although somewhat harsh, is entirely appropriate. Irrespective of what they are, what they can become, or what some users might assume them to be, we should treat all artifacts as mere tools and instruments. To its credit, this approach succeeds insofar as it reasserts and reconfirms the instrumental theory in the face of (perceived) challenges from a new kind of socially interactive and seemingly animate device. No matter how interactive, intelligent, or animated our AIs and robots become, they should be, now and forever, considered to be instruments or slaves in our service, nothing more. "We design, manufacture, own and operate robots," Bryson (2010, p. 65) writes. "They are entirely our responsibility. We determine their goals and behaviour, either directly or indirectly through specifying their intelligence, or even more indirectly by specifying how they acquire their own intelligence. But at the end of every indirection lies the fact that there would be no robots on this planet if it weren't for deliberate human decisions to create them."

There are, however, at least two problems with this proposal. First, it requires a kind of asceticism. Bryson's text issues what amounts to imperatives that take the form of social prohibitions directed to both designers and users. For designers, "thou shalt not create robots to be companions." For users, no matter how interactive or capable a robot is (or can become), "thou shalt not treat your robot as yourself." The validity and feasibility of these prohibitions, however, are challenged by actual data—not just anecdotal evidence gathered from the rather exceptional experiences of soldiers working with Packbots on the battlefield but numerous empirical studies of human/robot interaction that verify the media equation initially proposed by Reeves and Nass. In two recent studies (Rosenthal-von der Pütten et al. 2013 and Suzuki et al. 2015), for instance, researchers found that human users empathized with what appeared to be robot suffering even when they had prior experience with the device and knew that it was "just a machine." To put it in a rather crude vernacular form: Even when our head tells us it's just a robot, our heart cannot help but feel for it.[5]

Second, this way of thinking requires that we institute of a class of instrumental servant or slave. The problem here is not what one might think, namely, how the robot-slave might feel about its subjugation. The problem is with us and the effect this kind of institutionalized slavery could have on human individuals and communities. As de Tocqueville (2004) observed, slavery was not just a problem for the slave, it also had deleterious effects on the master and his social institutions. Clearly Bryson's use of the term "slave" is provocative and morally charged, and it would be impetuous to simply presume that this proposal for a kind of "slavery 2.0" would be the same or even substantially similar to what had occurred (and is still unfortunately occurring) with human bondage. But, and by the same token, we should also not dismiss or fail to take into account the documented evidence and historical data concerning slave-owning societies and how institutionalized forms of slavery affect things.

!S1 S2

The final modality also appears to support the independence and asymmetry of the two statements, but it does so by denying the first and affirming the second. In this case, which is something proposed and developed by Darling (2012, 2016), social robots, at least in term of the currently available technology, cannot have rights. They do not, at least at this particular point in time, possess the necessary capabilities or properties to be considered full moral and legal persons. Despite this fact, there is, Darling asserts, something qualitatively different about the way we encounter and perceive social robots. "Looking at state of the art technology, our robots are nowhere close to the intelligence and complexity of humans or animals, nor will they reach this stage in the near future. And yet, while it seems far-fetched for a robot's legal status to differ from that of a toaster, there is already a notable difference in how we interact with certain types of robotic objects" (Darling 2012, p. 1). This occurs, Darling continues, principally due to our tendencies to anthropomorphize things by projecting into them cognitive capabilities, emotions, and motivations that do not necessarily exist. Socially interactive robots, in particular, are intentionally designed to leverage and manipulate this proclivity. "Social robots," Darling (2012, p. 1) explains, "play off of this tendency by mimicking cues that we automatically associate with certain states of mind or feelings. Even in today's primitive form, this can elicit emotional reactions from people that are similar, for instance, to how we react to animals and to each other." And it is this emotional reaction that necessitates obligations in the face of social robots. "Given that many people already feel strongly about state-of-the-art social robot 'abuse,' it may soon become more widely perceived as out of line with our social values to treat robotic companions in a way that we would not treat our pets" (Darling 2012, p. 1).[6]

The obvious advantage to this way of thinking is that it is able to scale to recent technological developments in social robotics and the apparent changes they produced our in moral intuitions. Even if social robots cannot be moral subjects strictly speaking (at least not yet), there is something about this kind of machine that looks and feels different. According to Darling (2016, p. 213), it is because we "perceive robots differently than we do other objects," that one should consider extending some level of legal protections to the latter but not the former. This conclusion is consistent with Hume's thesis. If "ought" cannot be derived from "is," then axiological decisions concerning moral value are little more than sentiments based on how we feel about something at a particular time. Darling mobilizes a version of this moral sentimentalism with respect to social robots: "Violent behavior toward robotic objects *feels* wrong to many of us, even if we know that the abused object does not experience anything" (Darling 2016, p. 223). Consequently, and to its credit, Darling's proposal, unlike Bryson's "slavery 2.0" argument, tries to accommodate and work with rather than against recent and empirically documented experiences with social robots.

[5] Additional empirical evidence supporting this aspect of HRI (human robot interaction) has been tested and reported in the work of Bartneck and Hu (2008) and Bartneck et al. (2007).

[6] On the question of "robot abuse" see De Angeli et al. (2005), Bartneck and Hu (2008), Whitby (2008), and Nourbakhsh (2013, pp. 56–58).

There are, however, a number of complications with this approach. First, basing decisions concerning moral standing on individual perceptions and sentiment can be criticized for being capricious and inconsistent. "Feelings," Kant (1983, p. 442) writes in response to this kind of moral sentimentalism, "naturally differ from one another by an infinity of degrees, so that feelings are not capable of providing a uniform measure of good and evil; furthermore, they do so even though one man cannot by his feeling judge validly at all for other men." Additionally, because sentiment is a matter of individual experience, it remains uncertain as to whose perceptions actually matter or make the difference? Who, for instance, is included (and who is excluded) from the collective first person "we" that Darling operationalizes in and makes the subject of her proposal? In other words, whose sentiments count when it comes to decisions concerning the extension of moral and legal rights to others, and do all sentiments have the same status and value when compared to each other?

Second, despite the fact that Darling's proposal appears to uphold the Humean thesis, differentiating what ought to be from what is, it still proceeds by inferring "ought" from "is," or at least from what appears to be. According to Darling (2016, p. 214), everything depends on "our well-documented inclination" to anthropomorphize things. "People are prone," she argues, "to anthropomorphism; that is, we project our own inherent qualities onto other entities to make them seem more human-like"—qualities like emotions, intelligence, sentence, etc. Even though these capabilities do not (for now at least) really exist in the mechanism, we project them onto the robot in such a way that we then perceive them to be something that we presume actually belongs to the robot. By focusing on this anthropomorphic operation, Darling mobilizes and deploys a well-known Kantian distinction. What ultimately matters, according to her argument, is not what the robot actually is "in and of itself." What makes the difference is how the mechanism comes to be perceived. It is, in other words, the way the robot appears to us that determines how it comes to be treated. Although this change in perspective represents a shift from a kind of naïve empiricism to a more sophisticated phenomenological formulation (at least in the Kantian sense of the word), it still derives "ought," specifically how something ought to be treated, from what it appears to be.

Finally, because what ultimately matters is how "we" see things, this proposal remains thoroughly anthropocentric and instrumentalizes others. According to Darling, the principal reason we need to consider extending legal rights to others, like social robots, is for *our* sake. This follows the well-known Kantian argument for restricting animal abuse, and Darling endorses this formulation without any critical hesitation whatsoever: "The Kantian philosophical argument for preventing cruelty to animals is that our actions toward non-humans reflect our morality—if we treat animals in inhumane ways, we become inhumane persons. This logically extends to the treatment of robotic companions" (Darling 2016, pp. 227–228). This way of thinking, although potentially expedient for developing and justifying new forms of legal protections, renders the inclusion of previously excluded others less than altruistic; it transforms animals and robot companions into nothing more than instruments of human self-interest. The rights of others, in other words, is not about them; it is all about us.

Thinking otherwise

Although each modality has its advantages, none of the four provide what would be considered a definitive case either for or against robot rights. At this point, we can obviously continue to develop arguments and accumulate evidence supporting one or the other. But this effort, although entirely reasonable and justified, will simply elaborate what has formulated and will not necessarily advance the debate much further than what has already been achieved. In order to get some new perspective on the issue, we can (and perhaps should) try something different. This alternative, which could be called following Gunkel (2007) *thinking otherwise*, does not argue either for or against the is-ought inference but takes aim at and deconstructs this conceptual opposition. And it does so by deliberately flipping the Humean script, considering not "how ought may be derived from is" but rather "how is is only able to be derived from ought."

This is precisely the innovation introduced and developed by Emmanuel Levinas who, in direct opposition to the usual way of thinking, asserts that ethics precedes ontology. In other words, it is the axiological aspect, the "ought" dimension, that comes first, in terms of both temporal sequence and status, and then the ontological aspects follows from this decision.[7] This is a deliberate provocation that cuts across

[7] Although it is beyond the scope of this essay, it would be worth comparing the philosophical innovations of Emmanuel Levinas to the efforts of Knud Ejler Løgstrup (1997). In their "Introduction" to the English translation of Løgstrup's *The Ethical Demand*, Alasdair MacIntyre and Hans Fink (1997, xxxiii) offer the following comment: "Bauman in his *Postmodern Ethics* (Oxford: Blackwell, 1984)—a remarkable conspectus of a number of related postmodern standpoints—presents Løgstrup's work as having close affinities with Emmanuel Levinas. (Levinas was teaching at Strasbourg when Løgstrup was there in 1930, but there is no evidence that Løgstrup attended his Lectures.) Levinas, who was one of Husserl's students, was from the outset and has remained much closer to Husserlian phenomenology than Løgstrup ever was, often defining his own positions, even when they are antagonistic to Husserl's, in terms of their relationships to Husserl's. But, on some crucial issues, as Bauman's exposition makes clear, Levinas and Løgstrup are close." MacIntyer and Fink (1997, xxxiv) continue the comparison by noting important similarities that occur in the work of both Levinas and Løgstrup concerning the response to the Other, stressing that responsibility "is not derivable from or founded upon any universal rule or set of rights or determinate conception of the human good. For it is more fundamental in the moral life than any of these."

the grain of the philosophical tradition. As Floridi (2013, p. 116) correctly points out, in most moral theory, "what the entity is [the ontological question] determines the degree of moral value it enjoys, if any [the ethical question]." Levinas deliberately inverts and distorts this procedure. According to this way of thinking, we are first confronted with a mess of anonymous others who intrude on us and to whom we are obligated to respond even before we know anything at all about them. To use Hume's terminology—which will be a kind of translation insofar as Hume's philosophical vocabulary, and not just his language, is something that is foreign to Levinas's own formulations—we are first obligated to respond and then, after having made a response, what or who we responded to is able to be determined and identified. As Derrida (2005, p. 80) has characterized it, the crucial task in this alternative way of thinking moral consideration is to "reach a place *from* which the distinction between *who* and *what* comes to appear and become determined."

The advantage to this procedure is that it provides an entirely different method for responding to the challenge not just of social robots but of the way we have addressed and decided things in the face of this challenge. Following the contours of this Levinasian innovation, moral consideration is decided and conferred not on the basis of some pre-determined ontological criteria or capability (or lack thereof) but in the face of actual social relationships and interactions. "Moral consideration," as Coeckelbergh (2010, p. 214) describes it, "is no longer seen as being 'intrinsic' to the entity: instead it is seen as something that is 'extrinsic': it is attributed to entities within social relations and within a social context." In other words, as we encounter and interact with others—whether they be other human persons, an animal, the natural environment, or a social robot—this other entity is first and foremost situated in relationship to us. Consequently, the question of social and moral status does not necessarily depend on what the other is in its essence but on how she/he/it (and the pronoun that comes to be deployed in this situation is not immaterial) supervenes before us and how we decide, in "the face of the other" (to use Levinasian terminology), to respond. In this transaction, the "relations are prior to the things related" (Callicott 1989, p. 110), instituting what Gerdes (2015), following Coeckelbergh (2010), has called "a relational turn" in ethics.[8]

From the outset, this Levinasian influenced, relational ethic might appear to be similar to that developed by Kate Darling. Darling, as we have seen, also makes a case for extending moral and legal considerations to social robots irrespective of what they are or can be. The important difference, however, is anthropomorphism. For Darling, the reason we ought to treat robots well is because we perceive something of ourselves in them. Even if these traits and capabilities are not really there in the mechanism, we project them into or onto others—and other forms of otherness, like that represented by social robots—by way of anthropomorphism. For Darling, then, it is because the other looks and feels (to us) to be something like us that we are then obligated to extend to it some level of moral and legal consideration. For Levinas this anthropomorphic operation is *the* problem insofar as it reduces the other to a modality of the same—turning what is other into an alter-ego and mirrored projection of oneself. Levinas deliberately resists this gesture that already domesticates and even violates the alterity of the other. Ethics for Levinas is entirely otherwise: "The strangeness of the Other, his irreducibility to the I, to my thoughts and my possessions, is precisely accomplished as a calling into question of my spontaneity, as ethics" (Levinas 1969, p. 43). For Levinas, then, ethics is not based on "respect for others" but transpires in the face of the other and the interruption of *ipseity* that it produces. The principal moral gesture, therefore, is not the conferring or extending of rights to others as a kind of benevolent gesture or even an act of compassion but deciding how to respond to the Other who supervenes before me in such a way that always and already places my assumed rights and privilege in question.

This alternative configuration, therefore, does not so much answer or respond to the question with which we began as it alters the terms of the inquiry itself. When one asks "Can or should robots have rights?" the form of the question already makes an assumption, namely that rights are a kind of personal property or possession that an entity can have or should be bestowed with. Levinas does not inquire about rights nor does his moral theory attempt to respond to this form of questioning. In fact, the word "rights" is not in his philosophical vocabulary and does not appear as such in his published work. Consequently, the Levinasian question is directed and situated otherwise: "What does it take for something—another human person, an animal, a mere object, or a social robot—to supervene and be revealed as Other?" This *other question*—a question about others that is situated otherwise—comprises a more precise and properly altruistic inquiry. It is a mode of questioning that remains open, endlessly open, to others and other forms of otherness. For this reason, it deliberately interrupts and resists the imposition of power that Birch (1993, p. 39) finds operative in all forms of rights discourse: "The nub of the problem with granting or extending rights to others…is that it presupposes the existence and the maintenance of a position of power from which to do the granting." Whereas Darling is interested in "extending legal protection to social robots,"

[8] For a critical examination of the "relational-turn" applied to the problematic of animal rights philosophy, see Coeckelbergh and Gunkel's co-authored paper "Facing Animals: A Relational, Other-Oriented Approach to Moral Standing" (2014), the critical commentary provided by Piekarski (2016), and Coeckelbergh and Gunkel's (2016) reply to Piekarski's criticisms.

Levinas provides a way to question the assumptions and consequences involved in this very gesture. What we see in the face or the faceplate of the social robot, then, is not just a question concerning the rights of others—and other forms of socially significant otherness—but a challenge to this very way of asking about moral patiency.

Levinasian philosophy, therefore, has the potential to reorient the way we think about social robots and the question concerning rights. This alternative, however, still has at least one significant consequence that cannot and should not be ignored. Utilizing Levasian thought for the purposes of robophilosophy requires fighting against and struggling to break free from the gravitational pull of Levinas's own anthropocentric interpretations. Whatever the import of his unique contribution, "Other" in Levinas is still unapologetically human. Although he is not the first to identify it, Nealon (1998, p. 71) provides what is perhaps one of the most succinct descriptions of this problem: "In thematizing response solely in terms of the human face and voice, it would seem that Levinas leaves untouched the oldest and perhaps most sinister unexamined privilege of the same: *anthropos* [ἀνθρωπος] and only anthropos, has *logos* [λόγος]; and as such, *anthropos* responds not to the barbarous or the inanimate, but only to those who qualify for the privilege of 'humanity,' only those deemed to possess a face, only to those recognized to be living in the *logos*." For Levinas, as for many of those who follow in the wake of his influence, Other has been exclusively operationalized as another human subject. If, as Levinas argues, ethics precedes ontology, then in Levinas's own work anthropology and a certain brand of humanism still precede ethics.

This is not necessarily the only or even best possible outcome. In fact, Levinas can maintain this anthropocentrism only by turning "face" into a kind of ontological property and thereby undermining and even invalidating much of his own philosophical innovations. For others, like Matthew Calaraco, this is not and should not be the final word on the matter: "Although Levinas himself is for the most part unabashedly and dogmatically anthropocentric, the underlying logic of his thought permits no such anthropocentrism. When read rigorously, the logic of Levinas's account of ethics does not allow for either of these two claims. In fact…Levinas's ethical philosophy is, or at least should be, committed to a notion of universal ethical consideration, that is, an agnostic form of ethical consideration that has no a priori constraints or boundaries" (Calarco 2008, p. 55). In proposing this alternative reading, Calarco interprets Levinas against himself, arguing that the logic of Levinas's account is in fact richer and more radical than the limited interpretation the philosopher had initially provided for it. "If this is indeed the case," Calarco (2008, p. 55) concludes, "that is, if it is the case that we do not know where the face begins and ends, where moral considerability begins and ends, then we are obligated to proceed from the possibility that anything might take on a face. And we are further obligated to hold this possibility permanently open" (2008, p. 55). This means, of course, that we would be obligated to consider all kinds of others as Other, including other human persons, animals, the natural environment, artifacts, technologies, and robots. An "altruism" that tries to limit in advance who can or should be Other would not be, strictly speaking, altruistic.

Conclusions

Although the question concerning machine moral patiency has been a minor thread in robot ethics, there are good reasons to pursue the inquiry. Whether intended to do so or not, social robots effectively challenge unquestioned assumptions about technology and require that we make a decision—even if it is a decision not to decide—concerning the position and status of these socially interactive mechanisms. This essay has engaged this material by asking about rights, specifically "Can and should social robots have rights?" The question is formulated in terms of an historically important philosophical distinction—the is-ought problem—and it has generated four different kinds of responses, which are currently available in the existing literature. Rather than continue to pursue one or even a combination of these available modalities, I have proposed an alternative that addresses things otherwise. This *other question*—a question that is not just different but also open to difference—capitalizes on the philosophical innovations of Emmanuel Levinas and endeavors not so much to answer the question "Can and should robots have rights?" but to reformulate the way one asks about moral patiency in the first place. This outcome is consistent with the two vectors of robophilosophy: to apply philosophical thinking to the unique challenges and opportunities of social robots and to permit the challenge confronted in the face of social robots to question and reconfigure philosophical thought itself.

Acknowledgements This essay was written for and presented at Robophilosophy 2016, which took place at Aarhus University in Aarhus, Denmark (17–21 October 2016). My sincere thanks to Johanna Seibt, Marco Nørskov, and the programming committee for the kind invitation to participate in this event and to the other participants who contributed insightful questions and comments that have (I hope) been addressed in this published version.

References

Anderson, M., & Anderson, S. L. (2011). *Machine ethics*. Cambridge: Cambridge University Press.

Asaro, P. (2006). What should we want from a robot ethic? *International Review of Information Ethics, 6* (12), 9–16. http://www.i-r-i-e.net/inhalt/006/006_full.pdf.

Bartneck, C., van der Hoek, M., Mubin, O., & Mahmud, A. A. (2007). Daisy, daisy, give me your answer do!—Switching off a robot. In *Proceedings of the 2nd ACM/IEEE international conference on human-robot interaction* (pp. 217–222). doi:10.1145/1228716.1228746.

Bartnek, C., & Hu, J. (2008). Exploring the abuse of robots. *Interaction Studies, 9*(3), 415–433. doi:10.1075/is.9.3.04bar.

Bauman, Z. (1984). *Postmodern ethics*. Oxford: Blackwell.

Bentham, J. (1780). *An introduction to the principles of morals and legislation* (J. H. Burns & H. L. Hart, Ed.). Oxford: Oxford University Press, 2005.

Birch, T. (1993). Moral considerability and universal consideration. *Environmental Ethics, 15*, 313–332.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. New York: Oxford University Press.

Breazeal, C. L. (2002). *Designing sociable robots*. Cambridge, MA: MIT Press.

Bryson, J. (2010). Robots should be slaves. In Y. Wilks (Ed.), *Close engagements with artificial companions: Key social, psychological, ethical and design issues* (pp. 63–74). Amsterdam: John Benjamins.

Bryson, J. (2016). Patiency is not a virtue: AI and the design of ethical systems. AAAI Spring Symposium Series. Ethical and Moral Considerations in Non-Human Agents. http://www.aaai.org/ocs/index.php/SSS/SSS16/paper/view/12686.

Calarco, M. (2008). *Zoographies: The question of the animal from Heidegger to Derrida*. New York: Columbia University Press.

Callicott, J. B. (1989). *In defense of the land ethic: Essays in environmental philosophy*. Albany, NY: State University of New York Press.

Carpenter, J. (2015). *Culture and human-robot interaction in militarized spaces: A war story*. New York: Ashgate.

Coeckelbergh, M. (2010). Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology, 12*(3), 209–221. doi:10.1007/s10676-010-9235-5.

Coeckelbergh, M. (2012). *Growing moral relations: Critique of moral status ascription*. New York: Palgrave MacMillan.

Coeckelbergh, M., & Gunkel, D. J. (2014). Facing animals: A relational, other-oriented approach to moral standing. *Journal of Agricultural & Environmental Ethics, 27*(5), 715–733. doi:10.1007/s10806-013-9486-3.

Coeckelbergh, M., & Gunkel, D. J. (2016). Response to "The Problem of the Question About Animal Ethics" by Michal Piekarski. *Journal of Agricultural and Environmental Ethics, 29*(4), 717–721. doi:10.1007/s10806-016-9627-6.

Committee on Legal Affairs. (2016). Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics. European Parliament. http://www.europarl.europa.eu/ sides/getDoc.do?pubRef=-//EP//NONSGML%2BCOMPARL%2BPE-582.443%2B01%2BDOC%2BPDF%2BV0//EN.

Darling, K. (2012). Extending legal protection to social robots. *IEEE Spectrum*. http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/extending-legal-protection-to-social-robots.

Darling, K. (2016). Extending legal protection to social robots: The effects of anthropomophism, empathy, and violent behavior toward robotic objects. In R. Calo, A. M. Froomkin & I. Kerr (Eds.), *Robot law* (pp. 213–231). Northampton, MA: Edward Elgar.

De Angeli, A., Brahnam, S., & Wallis, P. (2005). Abuse: The dark side of human–computer interaction. *Interact* 2005. http://www.agentabuse.org/.

De Tocqueville, A. (2004). *Democracy in America* (A. Goldhammer, Trans.). New York: Penguin.

Dennett, D. C. (1998). *Brainstorms: Philosophical essays on mind and psychology*. Cambridge, MA: MIT Press.

Derrida, J. (2005). *Paper Machine* (R. Bowlby, Trans.). Stanford, CA: Stanford University Press.

Derrida, J. (2008). *The animal that therefore I am* (M.-L. Mallet, Ed., D. Wills, Trans.). New York: Fordham University Press.

Feenberg, A. (1991). *Critical theory of technology*. Oxford: Oxford University Press.

Floridi, L. (2013). *The ethics of information*. Oxford: Oxford University Press.

Garreau, J. (2007). Bots on the ground: In the field of battle (or even above it), robots are a soldier's best friend. *Washington Post*, 6 May. http://www.washingtonpost.com/wp-dyn/content/article/2007/05/05/AR2007050501009.html.

Gerdes, A. (2015). The issue of moral consideration in robot ethics. *ACM SIGCAS Computers & Society, 45*(3), 274–280. doi:10.1145/2874239.2874278.

Goertzel, B. (2002). Thoughts on AI morality. *Dynamical Psychology: An International, Interdisciplinary Journal of Complex Mental Processes*. http://www.goertzel.org/dynapsyc/2002/AIMorality.htm.

Gunkel, D. J. (2007). *Thinking otherwise: Philosophy, communication, technology*. West Lafayette, IN: Purdue University Press.

Gunkel, D. J. (2012). *The machine question: Critical perspectives on AI, robots, and ethics*. Cambridge, MA: MIT Press.

Güzeldere, G. (1997). The many faces of consciousness: A field guide. In N. Block, O. Flanagan & G. Güzeldere (Eds.), *The nature of consciousness: Philosophical debates* (pp. 1–68). Cambridge, MA: MIT Press.

Hall, J. S. (2001). Ethics for machines. KurzweilAI.net, July 5. http://www.kurzweilai.net/ethics-for-machines.

Haraway, D. J. (2008). *When species meet*. Minneapolis, MN: University of Minnesota Press.

Heidegger, M. (1977). *The question concerning technology and other essays* (W. Lovitt, Trans.). New York: Harper & Row.

Hudson, W. D. (1969). *The is/ought question: A collection of papers on the central problem in moral philosophy*. London: Macmillan.

Hume, D. (1980). *A treatise of human nature*. New York: Oxford University Press.

Johnson, D. G. (2006). Computer systems: Moral entities but not moral agents. *Ethics and Information Technology, 8*(4), 195–204. doi:10.1007/s10676-006-9111-5.

Kant, I. (1983). *Grounding for the metaphysics of morals*. Indianapolis, IN: Hackett Publishing.

Kurzweil, R. (2005). *The singularity is near: When humans transcend biology*. New York: Viking.

Levinas, E. (1969). *Totality and infinity: An essay on exteriority* (A. Lingis, Trans.). Pittsburgh, PA: Duquesne University.

Levy, D. (2005). *Robots unlimited: Life in a virtual age*. Boca Raton, FL: CRC Press.

Levy, D. (2009). The ethical treatment of artificially conscious robots. *International Journal of Social Robotics, 1*(3), 209–216. doi:10.1007/s12369-009-0022-6.

Løgstrup, K. E. (1997). *The ethical demand* (H. Fink & A. MacIntyre, Trans.). Notre Dame: University of Notre Dame Press.

Lyotard, J. F. (1984). *The postmodern condition: A report on knowledge* (G. Bennington & B. Massumi, Trans.). Minneapolis, MN: University of Minnesota Press.

MacIntyre, A., & Fink, H. (1997). Introduction. In K. E. Løgstrup (Ed.), *The ethical demand*. Notre Dame: University of Notre Dame Press.

Marx, K. (1977). *Capital: A critique of political economy* (B. Fowkes, Trans.). New York: Vintage Books.
Nealon, J. (1998). *Alterity Politics: Ethics and performative subjectivity*. Durham, NC: Duke University Press.
Nourbakhsh, I. (2013). *Robot futures*. Cambridge: MIT Press.
Piekarski, M. (2016). The problem of the question about animal ethics: Discussion with Mark Coeckelbergh and David Gunkel. *Journal of Agricultural and Environmental Ethics, 29*(4), 705–715. doi:10.1007/s10806-016-9626-7.
Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge: Cambridge University Press.
Rosenthal-von der Pütten, A. M., Krämer, N. C., Hoffmann, L., Sobieraj, S., & Eimler, S. C. (2013). An experimental study on emotional reactions towards a robot. *International Journal of Social Robotics, 5*(1), 17–34. doi:10.1007/s12369-012-0173-8.
Schurz, G. (1997). *The is-ought problem: An investigation in philosophical logic*. Dordrecht: Springer.
Searle, J. (1964). How to derive "ought" from "is". *The Philosophical Review, 73*(1), 43–58.
Seibt, J., Nørskov, M., & Andersen, S. S. (2016). *What social robots can and should do: Proceedings of robophilosophy 2016*. Amsterdam: IOS Press.
Singer, P., & Sagan, A. (2009). When robots have feelings. *The Guardian*. https://www.theguardian.com/commentisfree/2009/dec/14/rage-against-machines-robots.
Singer, P. W. (2009). *Wired for war: The robotics revolution and conflict in the twenty-first century*. New York: Penguin Books.
Sparrow, R. (2004). The turing triage test. *Ethics and Information Technology, 6*(4), 203–213. doi:10.1007/s10676-004-6491-2.
Suzuki, Y., Galli, L., Ikeda, A., Itakura, S. & Kitazaki, M. (2015). Measuring empathy for human and robot hand pain using electroencephalography. *Scientific Reports, 5*. doi:10.1038/srep15924.
Turkle, S. (2012). *Alone together: Why we expect more from technology and less from each other*. New York: Basic Books.
Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford: Oxford University Press.
Velmans, M. (2000). *Understanding consciousness*. London: Routledge.
Wallach, W., & Allen, C. (2009). *Moral machines: Teaching robots right from wrong*. Oxford: Oxford University Press.
Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*. San Francisco: W. H. Freeman.
Whitby, B. (2008). Sometimes it's hard to be a robot: A call for action on the ethics of abusing artificial agents. *Interacting with Computers, 20*(3), 326–333. doi:10.1016/j.intcom.2008.02.002.