Philos. Technol. (2014) 27:113–132
DOI 10.1007/s13347-013-0121-z

SPECIAL ISSUE

# A Vindication of the Rights of Machines

**David J. Gunkel**

Received: 3 January 2013 / Accepted: 8 July 2013 / Published online: 30 July 2013


**Abstract** This essay responds to the machine question in the affirmative, arguing that artifacts, like robots, AI, and other autonomous systems, can no longer be legitimately excluded from moral consideration. The demonstration of this thesis proceeds in four parts or movements. The first and second parts approach the subject by investigating the two constitutive components of the ethical relationship—moral agency and patiency. In the process, they each demonstrate failure. This occurs not because the machine is somehow unable to achieve what is considered necessary and sufficient to be a moral agent or patient but because the characterization of agency and patiency already fail to accommodate others. The third and fourth parts respond to this problem by considering two recent alternatives—the all-encompassing ontocentric approach of Luciano Floridi's information ethics and Emmanuel Levinas's eccentric ethics of otherness. Both alternatives, despite considerable promise to reconfigure the scope of moral thinking by addressing previously excluded others, like the machine, also fail but for other reasons. Consequently, the essay concludes not by accommodating the alterity of the machine to the requirements of moral philosophy but by questioning the systemic limitations of moral reasoning, requiring not just an extension of rights to machines, but a thorough examination of the way moral standing has been configured in the first place.

**Keywords** Artificial intelligence · Ethics · Machine ethics · Animal ethics · Information ethics

## 1 Introduction

One of the enduring concerns of moral philosophy is determining who or what is deserving of ethical consideration. Although initially limited to "other men," the practice of ethics has developed in such a way that it continually challenges its own restrictions and comes to encompass what had been previously excluded individuals and groups—foreigners, women, animals, and even the environment. "In the history

D. J. Gunkel (✉)
Department of Communication, Northern Illinois University, DeKalb, IL 60115, USA
e-mail: dgunkel@niu.edu

of the United States," Susan Leigh Anderson (2008, 480) has argued, "gradually more and more beings have been granted the same rights that others possessed and we've become a more ethical society as a result. Ethicists are currently struggling with the question of whether at least some higher animals should have rights, and the status of human fetuses has been debated as well. On the horizon looms the question of whether intelligent machines should have moral standing." The following responds to this final question—what we might call the "machine question" in ethics—in the affirmative, arguing that machines, like robots, AI, and other autonomous systems, can no longer and perhaps never really could be excluded from moral consideration. Toward that end, this paper advances another "vindication discourse," following in a tradition that begins with Mary Wollstonecraft's *A Vindication of the Rights of Men* (1790) succeeded two years later by *A Vindication of the Rights of Woman* and Thomas Taylor's intentionally sarcastic yet remarkably influential response *A Vindication of the Rights of Brutes*.[1]

Although informed by and following in the tradition of these vindication discourses, or what Peter Singer (1989, 148) has also called a "liberation movement," the argument presented here will employ something of an unexpected approach and procedure. Arguments for the vindication of the rights of previously excluded others typically proceed by (a) defining or characterizing the criteria for moral considerability or what Thomas Birch (1993, 315) calls the conditions for membership in "the club of consideranda," and (b) demonstrating that some previously excluded entity or group of entities are in fact capable of achieving a threshold level for inclusion in this community of moral subjects. "The question of considerability has been cast," as Birch (1993, 314) explains, "and is still widely understood, in terms of a need for necessary and sufficient conditions which mandate practical respect for whomever or what ever fulfills them." The vindication of the rights of machines, however, will proceed otherwise. Instead of demonstrating that machines or at least one representative machine is able to achieve the necessary and sufficient conditions for moral standing (however that might come to be defined, characterized, and justified) the following both contests this procedure and demonstrates the opposite, showing how the very criteria that have been used to decide the question of moral considerability necessarily fail in the first place. Consequently, the vindication of the rights of machines will not, as one might have initially expected, concern some recent or future success in technology nor will it entail a description of or demonstration with a particular artifact; it will instead investigate fundamental failures in the procedures of moral philosophy itself—failures that render exclusion of the machine both questionable and morally suspect.

## 2 Moral Agency

Questions concerning moral standing typically begin by addressing agency. The decision to begin with this subject is not accidental, provisional, or capricious. It is dictated and prescribed by the history of moral philosophy, which has traditionally privileged agency and the figure of the moral agent in both theory and practice. As

[1] What is presented here in the form of a "vindication discourse" is an abbreviated version of an argument that is developed in greater detail and analytical depth in Gunkel (2012).

Luciano Floridi explains, moral philosophy, from the time of the ancient Greeks through the modern era and beyond, has been almost exclusively agent-oriented. "Virtue ethics, and Greek philosophy more generally," Floridi (1999, 41) writes, "concentrates its attention on the moral nature and development of the individual agent who performs the action. It can therefore be properly described as an agent-oriented, 'subjective ethics.'" Modern developments, although shifting the focus somewhat, retain this particular agent-oriented approach. "Developed in a world profoundly different from the small, non-Christian Athens, Utilitarianism, or more generally Consequentialism, Contractualism and Deontologism are the three most well-known theories that concentrate on the moral nature and value of the actions performed by the agent" (Floridi 1999, 41). Although shifting emphasis from the "moral nature and development of the individual agent" to the "moral nature and value" of his or her actions, western philosophy has been, with few exceptions (which we will get to shortly), organized and developed as an agent-oriented endeavor.

When considered from the perspective of the agent, ethics inevitably and unavoidably makes exclusive decisions about who is to be included in the community of moral subjects and what can be excluded from consideration. The choice of words here is not accidental. As Jacques Derrida (2005, 80) points out everything turns on and is decided by the difference that separates the "who" from the "what." Moral agency has been customarily restricted to those entities who call themselves and each other "man"—those beings who already give themselves the right to be considered someone who counts as opposed to something that does not. But who counts—who, in effect, gets to be situated under the term "who"—has never been entirely settled, and the historical development of moral philosophy can be interpreted as a progressive unfolding, where what had once been excluded (i.e., women, slaves, people of color, etc.) have slowly and not without considerable struggle and resistance been granted access to the gated community of moral agents and have thereby also come to be someone who counts.

Despite this progress, which is, depending on how one looks at it, either remarkable or insufferably protracted, there remain additional exclusions, most notably non-human animals and machines. Machines in particular have been understood to be mere artifacts that are designed, produced, and employed by human agents for human specified ends. This instrumentalist and anthropocentric understanding has achieved a remarkable level of acceptance and standardization, as is evident by the fact that it has remained in place and largely unchallenged from ancient to postmodern times—from at least Plato's *Phaedrus* to Jean-François Lyotard's *The Postmodern Condition*. Beginning with the animal rights movement, however, there has been considerable pressure to reconsider the ontological assumptions and moral consequences of this legacy of human exceptionalism.

Extending consideration to these other previously marginalized subjects has required a significant reworking of the concept of moral agency, one that is not dependent on genetic make-up, species identification, or some other spurious criteria. As Singer (1999, 87) describes it, "the biological facts upon which the boundary of our species is drawn do not have moral significance," and to decide questions of moral agency on this ground "would put us in the same position as racists who give preference to those who are members of their race." For this reason, the question of moral agency has come to be disengaged from identification with the human being

and is instead often referred to and made dependent upon the generic concept of "personhood." "There appears," G. E. Scott (1990, 7) writes, "to be more unanimity as regards the claim that in order for an individual to be a moral agent s/he must possess the relevant features of a person; or, in other words, that being a person is a necessary, if not sufficient, condition for being a moral agent." Corporations, for example, are artificial entities that are obviously otherwise than human, yet they are considered legal persons, having rights and responsibilities that are recognized and protected by both national and international law (French 1979). As promising as this "personist" innovation is, "the category of the person," to reuse terminology borrowed from Marcel Mauss (1985), is by no means settled and clearly defined. There is, in fact, little or no agreement concerning what makes someone or something a person and the literature on this subject is littered with different formulations and often incompatible criteria. "One might well hope," Daniel Dennett (1998, 267) writes, "that such an important concept, applied and denied so confidently, would have clearly formulatable necessary and sufficient conditions for ascription, but if it does, we have not yet discovered them. In the end there may be none to discover. In the end we may come to realize that the concept person is incoherent and obsolete."

In an effort to contend with, if not resolve this problem, researchers often focus on the one "person making" quality that appears on most, if not all, the lists of "personal properties," whether they include just a couple simple elements (Singer 1999, 87) or involve numerous "interactive capacities" (Smith 2010, 74), and that already has traction with practitioners and theorists—consciousness. "Without consciousness," John Locke (1996, 146) argued, "there is no person." Or as Kenneth Einar Himma (2009, 19) articulates it, "moral agency presupposes consciousness…and that the very concept of agency presupposes that agents are conscious." Formulated in this fashion, moral agency is something that is decided and made dependent on a prior determination of consciousness. If, for example, an animal or a machine can in fact be shown to possess "consciousness," then that entity would, on this account, need to be considered a legitimate moral agent. And not surprisingly, there has been considerable effort in the fields of philosophy, AI, and robotics to address the question of machine moral agency by targeting and examining the question and possibility (or impossibility) of machine consciousness.

This seemingly rational approach, however, runs into considerable complications. On the one hand, we do not, it seems, have any widely accepted characterization of "consciousness." The problem, then, is that consciousness, although crucial for deciding who is and who is not a moral agent, is itself a term that is ultimately undecided and considerably equivocal. "The term," as Max Velmans (2000, 5) points out, "means many different things to many different people, and no universally agreed core meaning exists." 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. And to make matters worse, 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," Güven Güzeldere (1997, 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."

On the other hand, even if it were possible to define consciousness or come to some tentative agreement concerning its necessary and sufficient conditions, we still lack any credible and certain way to determine its actual presence in another. Because consciousness is a property attributed to "other minds," its presence or lack thereof requires access to something that is and remains fundamentally inaccessible. "How does one determine," as Paul Churchland (1999, 67) famously characterized it, "whether something other than oneself—an alien creature, a sophisticated robot, a socially active computer, or even another human—is really a thinking, feeling, conscious being; rather than, for example, an unconscious automaton whose behavior arises from something other than genuine mental states?" And the available solutions to this "other minds problem," from reworkings and modifications of the Turing Test to functionalist approaches that endeavor to work around this problem altogether (Wallach and Allen 2009), only make things more complicated and indeterminate. "There is," as Dennett (1998, 172) points out, "no proving that something that seems to have an inner life does in fact have one—if by 'proving' we understand, as we often do, the evincing of evidence that can be seen to establish by principles already agreed upon that something is the case." 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 empirical science, strictly speaking.[2] In the end, not only are these tests unable to demonstrate with any certitude whether animals, machines, or other entities are in fact conscious and therefore legitimate moral persons (or not), we are left doubting whether we can even say the same for other human beings. As Ray Kurzweil (2005, 380) candidly concludes, "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)."

The question of machine moral agency, therefore, turns out to be anything but simple or definitive. This is not, it is important to note, because machines are somehow unable to be moral agents. It is rather a product of the fact that the term "moral agent," for all its importance and argumentative expediency, has been and remains an ambiguous, indeterminate, and rather noisy concept. What the consideration of machine moral agency demonstrates, therefore, is something that may not have been anticipated or sought. What is discovered in the process of pursuing this line of inquiry is not a satisfactory answer to the question whether machines are able to be moral agents or not. In fact, that question remains open and unanswered. What has been ascertained is that the concept of moral agency is already vague and imprecise such that it is (if applied strictly and rigorously) uncertain whether we—whoever this "we" includes—are in fact moral agents.

What has been demonstrated, therefore, is that moral agency, the issue that had been assumed to be the "correct" place to begin, turns out to be inconclusive.

[2] Attempts to resolve this problem often take the form of a pseudo-science called physiognomy, which endeavors to infer an entity's internal states of mind from the observation of its external expressions and behavior.

Although this could be regarded as a "failure," it is a particularly instructive failing. What is learned from this failure—assuming we continue to use this obviously "negative" word—is that moral agency is not necessarily some property that can be definitively ascertained or discovered in others prior to and in advance of their moral consideration. Instead moral standing may be something like what Kay Foerst has called a dynamic and socially constructed "honorarium" (Benford and Malartre 2007, 165) that comes to be conferred and assigned to others in the process of our interactions and relationships with them. In this way then, "moral standing is not," as Mark Coeckelbergh (2012, 25) argues, "about the entity but about us and about the relation between us and the entity." But then the deciding factor will no longer be one of agency; it will be a matter of patiency.

## 3 Moral Patiency

Moral patiency looks at the ethical relationship from the other side. It is concerned not with determining the moral character of the agent or weighing the ethical significance of his/her/its actions but with the victim, recipient, or receiver of such action. This approach is, as Mane Hajdin (1994), Luciano Floridi (1999 and 2013), and others have recognized, a significant alteration in procedure and a "non-standard" way to approach the question of moral rights and responsibilities. The model for this kind of transaction can be found in the innovations of animal rights philosophy. Whereas agent-oriented ethics have been concerned with determining whether someone is or is not a legitimate moral subject with rights and responsibilities, animal rights philosophy begins with an entirely different question—"Can they suffer?" (Bentham 2005, 283). What distinguishes this particular mode of inquiry, as Derrida (2008, 28) points out, is that it asks not about an ability or power of the active agent (however that would come to be defined) but about a fundamental passivity—the patience of the patient, words that are derived from the Latin verb patior, which connotes "suffering." "Thus the question will not be to know whether animals are of the type *zoon logon echon* [ζωον λόγον έχον] whether they *can* speak or reason thanks to that *capacity* or that *attribute* of the *logos* [λόγος], the *can-have* of the *logos*, the aptitude for the *logos*. The *first* and *decisive* question would be rather to know whether animals *can suffer*" (Derrida 2008, 27).

This seemingly simple and direct question introduces what turns out to be a major shift in the basic structure and procedures of moral thinking. On the one hand, it challenges the anthropocentric tradition in ethics by questioning the often unexamined privilege human beings have granted themselves. In effect, it institutes something like a Copernican revolution in moral philosophy. Just as Copernicus challenged the geocentric model of the cosmos and in the process undermined many of the presumptions of human exceptionalism, animal rights philosophy contests the established Ptolemaic system of ethics, deposing the anthropocentric privilege that had traditionally organized the moral universe. On the other hand, the effect of this fundamental shift in focus means that the one time closed field of ethics can be opened up to other kinds of non-human animals. In other words, who counts as morally significant are not just other "men" but all kinds of entities that had previously been marginalized and situated outside the gates of the moral community. "If a being suffers," Singer (1975, 9) writes, "there can be no

moral justification for refusing to take that suffering into consideration. No matter what the nature of the being, the principle of equality requires that its suffering be counted equally with the like suffering of any other being."

Initially there seems to be good reasons and opportunities for extending this innovation to machines, or at least some species of machines (Gunkel 2007). This is because the animal and the machine, beginning with the work of René Descartes, share a common ontological status and position. For Descartes, the human being was considered the sole creature capable of rational thought—the one entity able to say, and be certain in its saying, *cogito ergo sum*. Following from this, he had concluded that other animals not only lacked reason but were nothing more than mindless automata that, like clockwork mechanisms, simply followed predetermined instructions programmed in the disposition of their various parts or organs. Conceptualized in this fashion, the animal and the machine, or what Descartes identified with the hybrid, hyphenated term *bête-machine*, were effectively indistinguishable and ontologically the same. "If any such machine," Descartes (1988, 44) wrote, "had the organs and outward shape of a monkey or of some other animal that lacks reason, we should have no means of knowing that they did not possess entirely the same nature as these animals."

Despite this fundamental and apparently irreducible similitude, only one of the pair has been considered a legitimate subject of moral concern. Even though the fate of the machine, from Descartes forward was intimately coupled with that of the animal, only the animal (and only some animals, at that) has qualified for any level of ethical consideration. And this exclusivity has been asserted and justified on the grounds that the machine, unlike the animal, does not experience either pleasure or pain. Steve Torrence (2008, 502) calls this "the organic view of ethical status" and demonstrates how philosophers have typically distinguished organic or biological organisms, either naturally occurring or synthetically developed, that are sentient and therefore legitimate subjects of moral consideration from what are termed "mere machines"—mechanisms that have no moral standing whatsoever. Although this conclusion appears to be rather reasonable and intuitive, it fails for a number of reasons.

First, it has been practically disputed by the construction of various mechanisms that now appear to suffer or at least provide external evidence of something that looks like pain. As Derrida (2008, 81) recognized, "Descartes already spoke, as if by chance, of a machine that simulates the living animal so well that it 'cries out that you are hurting it.'" This comment, which appears in a brief parenthetical aside in Descartes' *Discourse on Method*, had been deployed in the course of an argument that sought to differentiate human beings from the animal by associating the latter with mere mechanisms. But the comment can, in light of the procedures and protocols of animal ethics, be read otherwise. That is, if it were indeed possible to construct a machine that did exactly what Descartes had postulated, that is, "cry out that you are hurting it," would we not also be obligated to conclude that such a mechanism was capable of experiencing pain? This is, it is important to note, not just a theoretical point or speculative thought experiment. Engineers have, in fact, constructed mechanisms that synthesize believable emotional responses (Bates 1994; Blumberg et al. 1996; Breazeal and Brooks 2004), like the dental-training robot Simroid "who" cries out in pain when students "hurt" it (Kokoro 2009), and designed systems capable of evidencing behaviors that look a lot like what we usually call pleasure and pain.

Second it can be contested on epistemologically grounds insofar as suffering or the experience of pain is still unable to get around or resolve the problem of other minds. How, for example, can one know that an animal or even another person actually suffers? How is it possible to access and evaluate the suffering that is experienced by another? "Modern philosophy," Matthew Calarco (2008, 119) writes, "true to its Cartesian and scientific aspirations, is interested in the indubitable rather than the undeniable. Philosophers want proof that animals actually suffer, that animals are aware of their suffering, and they require an argument for why animal suffering should count on equal par with human suffering." But such indubitable and certain knowledge, as explained by Marian S. Dawkins, appears to be unattainable:

> At first sight, 'suffering' and 'scientific' are not terms that can or should be considered together. When applied to ourselves, 'suffering' refers to the subjective experience of unpleasant emotions such as fear, pain and frustration that are private and known only to the person experiencing them. To use the term in relation to non-human animals, therefore, is to make the assumption that they too have subjective experiences that are private to them and therefore unknowable by us. 'Scientific' on the other hand, means the acquisition of knowledge through the testing of hypotheses using publicly observable events. The problem is that we know so little about human consciousness that we do not know what publicly observable events to look for in ourselves, let alone other species, to ascertain whether they are subjectively experiencing anything like our suffering. The scientific study of animal suffering would, therefore, seem to rest on an inherent contradiction: it requires the testing of the untestable (Dawkins 2008, 1).

Because suffering is understood to be a subjective and private affair, there is no way to know, with any certainty or credible empirical method, exactly how another entity experiences unpleasant sensations such as fear, pain, or frustration. For this reason, it appears that the suffering of another (especially an animal) remains fundamentally inaccessible and unknowable. As Singer (1975, 11) readily admits, "we cannot directly experience anyone else's pain, whether that 'anyone' is our best friend or a stray dog. Pain is a state of consciousness, a 'mental event,' and as such it can never be observed." The question of machine moral patiency, therefore, leads to an outcome that was not necessarily anticipated. The basic problem is not whether the question "can they suffer?" applies to machines but whether anything that appears to suffer—human, animal, plant, or machine—actually does so at all.

Third, and to make matters even more complicated, we may not even know what "pain" and "the experience of pain" is in the first place. This point is something that is taken up and demonstrated by Dennett's "Why You Can't Make a Computer That Feels Pain" (1998). In this provocatively titled essay, originally published decades before the debut of even a rudimentary working prototype, Dennett imagines trying to disprove the standard argument for human (and animal) exceptionalism "by actually writing a pain program, or designing a pain-feeling robot" (191). At the end of what turns out to be a rather protracted and detailed consideration of the problem, he concludes that we cannot, in fact, make a computer that feels pain. But the reason for drawing this conclusion does not derive from what one might expect, nor does it offer

any kind of support for the advocates of moral exceptionalism. According to Dennett, the reason you cannot make a computer that feels pain is not the result of some technological limitation with the mechanism or its programming. It is a product of the fact that we remain unable to decide what pain is in the first place. The best we are able to do, as Dennett illustrates, is account for the various "causes and effects of pain," but "pain itself does not appear" (218). What is demonstrated, therefore, is not that some workable concept of pain cannot come to be instantiated in the mechanism of a computer or a robot, either now or in the foreseeable future, but that the very concept of pain that would be instantiated is already arbitrary, inconclusive, and indeterminate. "There can," Dennett writes at the end of the essay, "be 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" (228). Although Bentham's question "Can they suffer?" may have radically reoriented the direction of moral philosophy, the fact remains that "pain" and "suffering" are just as nebulous and difficult to define and locate as the concepts they were intended to replace.

Finally, all this talk about the possibility of engineering pain or suffering in a machine entails its own particular moral dilemma. "If (ro)bots might one day be capable of experiencing pain and other affective states," Wallach and Allen (2009), 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 and should it be prohibited?" If it were in fact possible to construct a machine that "feels pain" (however defined and instantiated) in order to demonstrate the limits of moral patiency, 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 needs to be able to construct such a mechanism in order to demonstrate moral patiency and the moral standing of machines; but doing so would be, on that account, already to engage in an act that could potentially be considered immoral. Or to put it another way, the demonstration of machine moral patiency might itself be something that is quite painful for others.

Despite initial promises, we cannot, it seems, make a credible case for or against the moral standing of the machine by simply following the patient-oriented approach modeled by animal rights philosophy. In fact, trying to do so produces some rather unexpected results. In particular, extending these innovations does not provide definitive proof that the machine either can be or is not able to be a similarly constructed moral patient. Instead doing so demonstrates how the "animal question"—the question that had in effect revolutionized ethics in the later half of the 20$^{th}$ century—might already be misguided and prejudicial. Although it was not necessarily designed to work in this fashion, "A Vindication of the Rights of Machines" achieves something similar to what Thomas Taylor had wanted for his *A Vindication of the Rights of Brutes*. Taylor, who wrote and distributed this pamphlet under the protection of anonymity, originally composed the essay as a means by which to parody and undermine the arguments that had been advanced in Wollstonecraft's *A Vindication of the Rights of Woman*. Taylor's text, in other words, was initially offered as a kind of *reductio ad absurdum* designed to exhibit what he perceived to be the conceptual failings of Wollstonecraft's proto-

feminist manifesto. Following suit, "A Vindication of the Rights of Machines" appears to have the effect of questioning and even destabilizing what had been achieved with animal rights philosophy. But as was the case with the consideration of moral agency, this negative outcome is informative and telling. In particular, it indicates to what extent this apparent revolution in moral thinking is, for all its insight and promise, still beset with fundamental problems that proceed not so much from the ontological condition of these other, previously excluded entities but from systemic problems in the very structure and protocols of moral reasoning.

## 4 Information Ethics

One of the criticisms of animal rights philosophy is that this moral innovation, for all its promise to intervene in the anthropocentric tradition, remains an exclusive and exclusionary practice. "If dominant forms of ethical theory," Calarco (2008, 126) argues "—from Kantianism to care ethics to moral rights theory—are unwilling to make a place for animals within their scope of consideration, it is clear that emerging theories of ethics that are more open and expansive with regard to animals are able to develop their positions only by making other, equally serious kinds of exclusions…" Environmental and land ethics, for instance, have been critical of animal rights philosophy for including some sentient creatures in the community of moral patients while simultaneously excluding other kinds of animals, plants, and the other entities that comprise the natural environment. In response to this exclusivity, environmental ethicists have argued for a further expansion of the moral community to include these marginalized others, or the excluded other of the animal other.

Although these efforts effectively expand the community of legitimate moral patients to include those others who had been previously left out, environmental ethics has also (and not surprisingly) been criticized for instituting additional omissions. "Even bioethics and environmental ethics," Floridi (2013, 64) argues, "fail to achieve a level of complete universality and impartiality, because they are still biased against what is inanimate, lifeless, intangible, abstract, engineered, artificial, synthetic, hybrid, or merely possible. Even land ethics is biased against technology and artefacts, for example. From their perspective, only what is intuitively alive deserves to be considered as a proper centre of moral claims, no matter how minimal, so a whole universe escapes their attention." According to this line of reasoning, bioethics and environmental ethics represents something of an incomplete innovation in moral philosophy. They have, to their credit, successfully challenged the excluded other of animal rights philosophy by articulating a more universal form of ethics that not only shifts attention to the patient but also expands who or what qualifies for inclusion as a moral patient. At the same time, however, these innovations remain ethically biased insofar as they substitute a biocentrism for animocentrism and in the process continue to exclude other entities, specifically technology and other kinds of artifacts.

In response to this, Floridi endeavors to take the innovations introduced by bioethics and environmental ethics one step further. He adopts their patient-oriented approach but "lowers the condition that needs to be satisfied, in order to qualify as a centre of moral concern, to the minimal common factor shared by any entity" (Floridi 2013, 64) whether animate, inanimate, or otherwise. For Floridi this lowest common

denominator is informational and, for this reason, he gives his innovative proposal the name "Information Ethics" or IE. "IE is an ecological ethics that replaces biocentrism with ontocentrism. IE suggests that there is something even more elemental than life, namely being—that is, the existence and flourishing of all entities and their global environment—and something more fundamental than suffering, namely entropy, [which] here refers to any kind of destruction or corruption of informational objects, that is, any form of impoverishment of being including nothingness, to phrase it more metaphysically" (Floridi 2008, 47). Following the innovations of bio- and environmental ethics, Floridi expands the scope of moral philosophy by altering its focus and lowering the threshold for inclusion, or, to use Floridi's terminology, the level of abstraction (LoA). What makes someone or something a moral patient, deserving of some level of ethical consideration, is that it exists as a coherent body of information. Consequently, something can be said to be good, from an IE perspective, insofar as it respects and facilitates the informational welfare of a being and bad insofar as it causes diminishment, leading to an increase in information entropy. In fact, for IE, "fighting information entropy is the general moral law to be followed" (Floridi 2002, 300). This fundamental shift in focus provides for a moral theory that is more inclusive of others. "Unlike other non-standard ethics," Floridi (2013, 65) argues, "IE is more impartial and universal—or one may say less ethically biased—because it brings to ultimate completion the process of enlarging the concept of what may count as a centre of moral claims, which now includes every instance of information, no matter whether physically implemented or not."

Despite the fact that IE promises to bring "to ultimate completion" the patient-oriented innovation of bio-ethics, the proposal is not without its problems. First, in shifting emphasis from an agent-oriented to a patient-oriented ethics, IE (like animal rights philosophy and bio-ethics before it) simply inverts the two terms of a traditional binary opposition. If classic ethical thinking has been organized, for better or worse, by an interest in the character and/or actions of the agent at the expense of the patient, IE endeavors, following the example of previous innovations, to reorient things by placing emphasis on the other term. This maneuver is, quite literally, a revolutionary proposal, because it inverts or "turns over" the traditional arrangement. Inversion, however, is rarely in and by itself a satisfactory mode of intervention. As Nietzsche (1974), Heidegger (1962), Derrida (1978), and other poststructuralists have pointed out, the inversion of a binary opposition actually does little or nothing to disturb or to challenge the fundamental structure of the system in question (Gunkel 2007). In fact, inversion preserves and maintains the traditional structure, albeit in an inverted form. The effect of this on IE has been registered by Himma, who, in an assessment of Floridi's initial publications on the subject, demonstrates that a concern for the patient is really nothing more than the flip-side of good-old, agent-oriented, anthropocentric ethics. "To say that an entity X has moral standing (i.e., is a moral patient) is, at bottom, simply to say that it is possible for a moral agent to commit a wrong against X. Thus, X has moral standing if and only if (1) some moral agent has at least one duty regarding the treatment of X and (2) that duty is owed to X" (Himma 2004, 145). According to Himma's analysis, IE's patient-oriented ethics (or any patient-oriented ethics, for that matter) is not that different from traditional forms of agent-oriented ethics. It simply looks at the agent/patient couple from the other side and in doing so still operates on and according to the standard system. Although

instituting a revolutionary alteration in perspective, IE's patient-oriented ethics do not necessarily change the rules of the game.

Second, IE is not limited to simply turning things around. It also enlarges the scope of moral consideration by reducing the minimum requirements for inclusion. "IE holds," Floridi (2013, 68–69) argues, "that every informational entity, insofar as it is an expression of Being, has a dignity constituted by its mode of existence and essence, defined here as the collection of all the elementary proprieties that constitute it for what it is." Like previous innovations, IE is interested in expanding membership in the moral community so as to incorporate previously excluded others. But, unlike previous efforts, it is arguably more inclusive of others and other forms of otherness. IE, therefore, contests and seeks to replace both the exclusive anthropocentric and biocentric theories with an "ontocentric" one, which is, by comparison, much more inclusive and universal. In taking this approach, however, IE simply substitutes one form of centrism for another. Anthropocentrism, for example, situates the human at the center of moral concern and admits into consideration anyone who is able to meet the basic criteria of what has been decided to comprise the human. Animiocentrism focuses attention on the animal and extends consideration to any organism that meets the defining criteria of animality. Biocentrism goes one step further in the process of abstraction; it defines life as the common denominator and admits into consideration anything and everything that can be said to be alive. And ontocentrism completes the progression by incorporating into moral consideration anything that actually exists, had existed, or potentially exists.

All of these innovations, despite their differences in focus, employ a similar maneuver and logic. That is, they redefine the center of moral consideration in order to describe progressively larger circles that come to encompass a wider range of possible participants. Although there are and will continue to be considerable debates about what should define the center and who or what is or is not included, this debate is not the problem. The problem rests with the strategy itself. In taking a centrist approach, these different ethical theories (of which IE would presumably be the final and ultimate form) endeavor to identify what is essentially the same in a phenomenal diversity of different individuals. Consequently, they include others by effectively stripping away and reducing differences. This approach, although having the appearance of being increasingly more inclusive, effaces the unique alterity of others and turns them into more of the same. This is, according to Levinas (1969 and Levinas 1981) the defining gesture of philosophy and one that does considerable violence to others. "Western philosophy," Levinas (1969, 43) argues, "has most often been an ontology: a reduction of the other to the same by interposition of a middle or neutral term that ensures the comprehension of being" (Levinas 1969, 43). The issue, therefore, is not deciding which form of centrism is more or less inclusive of others; the difficulty rests with this strategy itself, which succeeds only by reducing difference and turning what is other into a modality of the same.

Finally, this metaphysical operation is never neutral, and its moral consequences have been identified by environmental ethicists like Thomas Birch, who finds any and all efforts to articulate criteria for "universal consideration" to be based on a fundamentally flawed assumption. According to Birch, these efforts at increasingly more inclusive inclusion always proceed by way of articulating some necessary and sufficient conditions, or qualifying characteristics, that must be met by an entity in

order to be incorporated into the community of legitimate moral subjects. In traditional forms of anthropocentric ethics, for example, it was the *anthropos* and the way it had been characterized (which it should be noted was always and already open to considerable social negotiation and redefinition), that provided the criteria for deciding who would be include in the moral community and what would not. The problem, Birch contends, is not necessarily with the criteria that are selected to make these decisions (although it is possible to argue that there have been better and worse formulations); the more fundamental problem is with the patient-oriented strategy and approach. "The institution of any practice of any criterion of moral considerablity," Birch (1993, 317) writes, "is an act of power over, and ultimately an act of violence" toward others. In other words, every criteria of moral inclusion, no matter how neutral, objective, or universal it appears, is an imposition of power insofar as it consists in the universalization of a particular value or set of values made by someone from particular position of power. "The nub of the problem with granting or extending rights to others," Birch (1995, 39) concludes, "a problem which becomes pronounced when nature is the intended beneficiary, is that it presupposes the existence and the maintenance of a position of power from which to do the granting." Even in the case of the relatively more inclusive and seemingly all-encompassing patient-oriented approach instituted by IE, someone has already been empowered to decide what particular criteria will be considered the necessary and sufficient conditions for inclusion in the class of "moral consideranda" (Birch 1993, 317).[3] The problem, then, is not only with the specific criteria that comes to be selected as the universal condition but also, and more so, the very act of universalization, which already empowers someone to make these decisions for others.

Although IE provides for a more complete and universal articulation of a patient-oriented ethics able to include others, including machines, this all-encompassing totalizing effect is simultaneously its greatest achievement and a critical problem. It is an achievement insofar as it carries through to completion the patient-oriented approach that begins to gain momentum with animal rights philosophy. IE promises, as Floridi (1999) describes it, to articulate an "ontocentric, patient-oriented, ecological macroethics" that includes everything, does not make other problematic exclusions, and is sufficiently universal, complete, and consistent. It is a problem insofar as this approach to greater inclusivity continues to deploy and support a strategy that is itself part and parcel of a "totalizing" (Levinas 1969) or "imperialist" (Birch 1993) program. The problem, then, is not which centrism one develops and patronizes or which criteria are determined to be more or less inclusive; the problem is with this approach itself. What is the matter with IE, therefore, is not the way Floridi develops this ultimate form of patient-oriented ethics, which has a good deal to commend it. The problem is with the patient-oriented methodology that it inherits, deploys, and leaves largely uninterogated. What is needed, therefore, is another approach, one that is not satisfied with being merely revolutionary in its innovations, one that does not continue to pursue a project of totalizing and potentially violent assimilation, and one that can respond to and take responsibility for what remains in excess of the entire

[3] Although it could be argued that Being is so general a criterion that it must escape this criticism, the fact of the matter is that Being is a concern of and for a particular being. In fact, Heidegger (1962, 32) famously defined the human being as that entity for whom Being is an issue: "Dasein is ontically distinctive in that it is ontological." Understood in this way, it is possible to conclude that IE is just another form of anthropocentric ethics insofar as its ontocentric focus is the defining condition of human Dasein.

conceptual field that has been delimited and defined by the binary pair of agent and patient. What is needed is some way of proceeding and thinking otherwise—a way that, in the context of and in response to IE's ontocentric ethics, would be "otherwise than being" (Levinas 1981).

## 5 Thinking Otherwise

When it comes to thinking otherwise, especially as it relates to the question concerning ethics, there is perhaps no philosopher better suited to the task than Emmanuel Levinas. Unlike a lot of what goes by the name of "moral philosophy," Levinasian thought does not rely on metaphysical generalizations, abstract formulas, or simple pieties. It is not only critical of the traditional tropes and traps of western ontology but proposes an "ethics of otherness" that deliberately resists and interrupts the metaphysical gesture par excellence, that is, the reduction of difference to the same. This radically different approach to thinking difference differently is not just a useful and expedient strategy. It is not, in other words, a mere gimmick. It constitutes a fundamental reorientation that effectively alters the rules of the game and the standard operating presumptions. In this way, "morality is," as Levinas (1969, 304) concludes, "not a branch of philosophy, but first philosophy." This fundamental reconfiguration, which puts ethics first in both sequence and status, permits Levinas to circumvent and deflect a lot of the difficulties that have traditionally tripped up moral thinking in general and efforts to address the moral status of the machine in particular.

First, for Levinas, the problems of other minds[4]—a difficulty, as we have seen, for both agent-oriented and patient-oriented approaches—is not some fundamental epistemological limitation that must be addressed and resolved prior to moral decision making but constitutes the very condition of the ethical relationship as an irreducible exposure to an other who always and already exceeds the boundaries of one's totalizing comprehension. Consequently Levinasian philosophy, instead of being derailed by the epistemological problem of other minds, immediately affirms and acknowledges it as the basic condition of possibility for ethics. Or as Richard Cohen (2001, 336) succinctly describes it in what could be a marketing slogan for Levinasian thought, "not 'other minds,' mind you, but the 'face' of the other, and the faces of all others." In this way, then, Levinas provides for a seemingly more attentive and empirically grounded approach to the problem of other minds insofar as he explicitly acknowledges and endeavors to respond to and take responsibility for the original and irreducible difference of others instead of getting involved with and playing all kinds of speculative (and unfortunately wrong-headed) head games. "The ethical relationship," Levinas (1987, 56) writes, "is not grafted on to an antecedent relationship of cognition; it is a foundation and not a superstructure…It is then more cognitive than cognition itself, and all objectivity must participate in it."

Second, and following from this, Levinas's concern with/for the Other (which is often capitalized like a proper name) will constitute neither an agent nor patient

[4] This analytic moniker is something that is not ever used by Levinas, who is arguably the most influential moral thinker in the continental tradition. The term, however, has been employed by a number of Levinas's Anglophone interpreters.

oriented ethics, but addresses itself to what is anterior to and remains in excess of this seemingly fundamental binary structure—the basic structure that, Floridi (1999, 41) asserts, constitutes the logical form of any and all action, whether morally loaded or not. Although Levinas's attention to and concern for others looks, from one perspective at least, to be a kind of "patient oriented" ethics that puts the interests and rights of the other before oneself, it is not and cannot be satisfied with simply endorsing one side of or conforming to the agent/patient couple. Unlike Floridi's IE, which advocates a patient-oriented ethics in opposition to the customary agent-oriented approaches that have maintained a controlling interest in the field, Levinas goes one step further, releasing what could be called a deconstruction[5] of the very conceptual order of agent and patient. This alternative, as Levinas (1981, 117) explains is located "on the hither side of the act-passivity alternative" and, for that reason, significantly reconfigures the standard terms and conditions. "For the condition for," Levinas (1981, 123) explains, "or the unconditionality of, the self does not begin in the auto-affection of a sovereign ego that would be, after the event, 'compassionate' for another. Quite the contrary: the uniqueness of the responsible ego is possible only in being obsessed by another, in the trauma suffered prior to any auto-identification, in an unrepresentable before." The self or the ego, as Levinas describes it, does not constitute some preexisting self-assured condition that is situated before and as the cause of the subsequent relationship with an other. It does not (yet) take the form of an active agent who is able to decide to extend him/herself to others in a deliberate act of compassion. Rather it becomes what it is as a byproduct of an uncontrolled and incomprehensible exposure to the face of the other that takes place prior to and in advance of any formulation of the self in terms of agency.

Likewise the Other is not comprehended as a patient who would be the recipient of the agent's actions and whose interests and rights would need to be identified, taken into account, and duly respected. Instead, the absolute and irreducible exposure to the Other is something that is anterior and exterior to these distinctions, not only remaining beyond the range of their conceptual grasp and regulation but also making possible and ordering the antagonistic structure that subsequently comes to characterize the difference that distinguishes the self from its others and the agent from the patient in the first place. In other words, for Levinas at least, prior determinations of agency and patiency do not first establish the terms and conditions of any and all possible encounters that the self might have with others and with other forms of otherness. It is the other way around. The Other first confronts, calls upon, and interrupts self-involvement and in the process determines the terms and conditions by which and in response to which the standard roles of moral agent and moral patient come to be articulated and assigned. Consequently, Levinas's philosophy is not what is typically understood as an ethics, a meta-ethics, a normative ethics, or even an applied ethics. It is, what John Llewelyn (1995, 4) has called a "proto-ethics" or what others have termed an "ethics of ethics." "It is true," Derrida explains, "that Ethics in Levinas's sense is an Ethics without law and without concept, which maintains its

[5] Employing the term "deconstruction" in this particular context is somewhat problematic. This is because deconstruction does not necessarily sit well with Levinas's own work. Levinas, both personally and intellectually, had a rather complex relationship with Jacques Derrida, the main proponent of what is often mislabeled "deconstructivism," and an even more complicated, if not contentious one with Martin Heidegger, the thinker who Derrida credits with having first introduced the concept and practice.

non-violent purity only before being determined as concepts and laws. This is not an objection: let us not forget that Levinas does not seek to propose laws or moral rules, does not seek to determine a morality, but rather the essence of the ethical relation in general. But as this determination does not offer itself as a theory of Ethics, in question, then, is an Ethics of Ethics" (Derrida 1978, 111). In comparison to the anthropocentrism of the standard, agent-oriented approach and the animocentric/biocentric/ontocentric efforts of the various non-standard, patient-oriented alternatives, we can say that Levinas proposes a truly eccentric philosophy that exceeds the orbit and conceptual grasp of both.

Despite the promise this innovation has for arranging a moral philosophy that is radically situated otherwise, Levinas's work remains committed to and is not able to escape the influence of anthropocentric privilege and human exceptionalism. Whatever the import of his unique contribution, "other" in Levinas is still and unapologetically human (Levinas 2003). Although he is not the first to identify it, Jeffrey Nealon provides what is perhaps one of the most succinct description of the 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 [άνθρωπος] 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*" Nealon (1998, 71). For Levinas, therefore, technological devices may have an interface, but they do not possess a face or confront us in a face-to-face encounter that would call for and would be called ethics. If Levinasian philosophy is to provide a way of thinking otherwise that is able to respond to and to take responsibility for these other forms of otherness (and not just machines but non-human animals as well), we will need to employ and interpret this innovation against and in excess of Levinas's own interpretation of it. We will need as Derrida (1978, 260) once wrote of Georges Bataille's exceedingly careful engagement with the thought of Hegel, to follow Levinas to the end, "to the point of agreeing with him against himself" and of wresting his discoveries from the limited interpretations that he had provided.

Such efforts at "radicalizing Levinas," as Atterton and Calarco (2010) refer to it, will take up and pursue Levinas's "ethics of otherness" in excess of and beyond the rather restricted formulations that he and his advocates and critics have typically provided. "Although Levinas himself is for the most part unabashedly and dogmatically anthropocentric," Calarco (2008, 55) writes, "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, 55).[6] This reworking of Levinasian philosophy promises to provide a much more inclusive articulation that is

[6] In stating this, we immediately run up against and need to confront the so-called problem of relativism—"the claim that no universally valid beliefs or values exist" (Ess 1996, 204). Although a complete response to this problem lies outside the scope of this particular essay, we should, at this point at least, recognize that "relativism" is not necessarily a pejorative term. For more on this issue see Scott (1967), Žižek (2006), and Gunkel (2012).

able to take other forms of otherness into account. And it is a compelling proposal. What is interesting about Calarco's argument (and the arguments offered by other Levinasian influenced thinkers, like Benso, 2000), however, is not the other forms of otherness that come to be included by way of this innovative reconfiguration of Levinasian thought, but what (unfortunately) gets left out in the process. According to the letter of Calarco's text the following entities could be given consideration: "'lower' animals, insects, dirt, hair, fingernails, and ecosystems" (Calarco 2008, 71). What is obviously missing from this list is anything that is not "natural," that is, any form of artifact or technology. Consequently, what gets left behind or left out by Calarco's "universal ethical consideration" are tools, technologies, and machines. Despite the fact that "universal consideration would entail being ethically attentive and open to the possibility that anything might take on face" (Calarco 2008, 73), machines appear to be the faceless constitutive exception. For this reason, "thinking otherwise," although clearly offering a compelling alternative to both agent and patient oriented ethics, still fails to respond to and take full responsibility for the machine.

## 6 Conclusion

"Every philosophy," Silvia Benso (2000, 136) writes in a comprehensive gesture that performs precisely what it seeks to address, "is a quest for wholeness." This objective, she argues, has been typically targeted in one of two ways. "Traditional Western thought has pursued wholeness by means of reduction, integration, systematization of all its parts. Totality has replaced wholeness, and the result is totalitarianism from which what is truly other escapes, revealing the deficiencies and fallacies of the attempted system." This is precisely the kind of violent philosophizing that Levinas (1969) identifies under the term "totality," and it includes the efforts of both standard agent-oriented and non-standard patient-oriented approaches up to and including information ethics. The alternative to these totalizing transactions is a philosophy that is oriented otherwise, like that proposed by Levinas. This other approach, however, "must do so by moving not from the same, but from the other, and not only the Other, but also the other of the Other, and, if that is the case, the other of the other of the Other. In this must, it must also be aware of the inescapable injustice embedded in any formulation of the other" (Benso 2000, 136). And this "injustice" is evident not only in Levinas's exclusive humanism but in the way that those who seek to redress this "humanism of the other" continue to ignore or marginalize the machine.

For these reasons, the question concerning machine moral standing does not end with a single definitive answer—a simple and direct "yes" or "no." But this inability is not, we can say following the argumentative strategy of Dennett's "Why you Cannot Make a Computer that Feels Pain" (1998), necessarily a product of some inherent or essential deficiency with the machine. Instead it is a result of the fact that moral agency, moral patiency, and those ethical theories that endeavor to think otherwise already deploy and rely on questionable constructions and logics. The vindication of the rights of machines, therefore, is not simply a matter of extending moral consideration to one more historically excluded other, which would, in effect,

leave the mechanisms of moral philosophy in place, fully operational, and unchallenged. Instead, the question concerning the "rights of machines" makes a fundamental claim on ethics, requiring us to rethink the system of moral considerability all the way down. This is, as Levinas (1981, 20) explains, the necessarily "interminable methodological movement of philosophy" that continually struggles against accepted practices in an effort to think otherwise—not just differently but in ways that are responsive to and responsible for others.

Consequently, this essay ends not as one might have expected. That is, by accumulating evidence or arguments in favor of permitting machines, or even one representative machine, entry into the community of moral subjects. Instead, it concludes with questions about ethics and the way moral philosophy has typically defined and decided moral standing. Although ending in this questionable way—in effect, responding to a question with a question—is commonly considered bad form, this is not necessarily the case. This is because questioning is a particularly philosophical enterprise. "I am," Dennett (1996, vii) writes, "a philosopher and not a scientist, and we philosophers are better at questions than answers. I haven't begun by insulting myself and my discipline, in spite of first appearances. Finding better questions to ask, and breaking old habits and traditions of asking, is a very difficult part of the grand human project of understanding ourselves and our world." The objective of the vindication of the rights of machines, therefore, has not been to answer the machine question with some definitive proof or preponderance of evidence, but to ask about the very means by which we have gone about trying to articulate and formulate this question. The issue, then, is not can the machine be a moral agent, a moral patient, or something else? Instead the question concerns how moral agency and patiency have been configured and how these configurations already accommodate and/or marginalize others. The vindication of the rights of machines, therefore, is not just one more version or iteration of an applied moral philosophy; it releases a thorough and profound challenge to what is called "ethics."

## References

Anderson, S. (2008). Asimov's 'Three Laws of Robotics' and Machine Metaethics. *AI & Society, 22*(4), 477–493.

Atterton, P., & Calarco, M. (2010). *Radicalizing Levinas*. Albany, NY: SUNY Press.

Bates, J. (1994). The Role of Emotion in Believable Agents. *Communications of the ACM, 37*, 122–125.

Benford, G., & Malartre, E. (2007). *Beyond Human: Living with Robots and Cyborgs*. New York: Tom Doherty.

Benso, S. (2000). *The Face of Things: A Different Side of Ethics*. Albany, NY: SUNY Press.

Bentham, J. (2005). *An Introduction to the Principles of Morals and Legislation. J. H. Burns and H. L. Hart (Eds.)*. Oxford: Oxford University Press.

Birch, T. (1993). Moral Considerability and Universal Consideration. *Environmental Ethics, 15*, 313–332.

Birch, T. (1995). The Incarnation of Wilderness: Wilderness Areas as Prisons. In M. Oelschlaeger (Ed.), *Postmodern Environmental Ethics* (pp. 137–162). Albany, NY: State University of New York Press.

Blumberg, B., Todd, P., & Maes, M. (1996). No Bad Dogs: Ethological Lessons for Learning. In *Proceedings of the 4th International Conference on Simulation of Adaptive Behavior* (pp. 295–304). Cambridge, MA: MIT Press.

Breazeal, C., & Brooks, R. (2004). Robot Emotion: A Functional Perspective. In J. M. Fellous & M. Arbib (Eds.), *Who Needs Emotions: The Brain Meets the Robot* (pp. 271–310). Oxford: Oxford University Press.

Calarco, M. (2008). *Zoographies: The Question of the Animal from Heidegger to Derrida*. New York: Columbia University Press.
Churchland, P. M. (1999). *Matter and Consciousness*. Cambridge, MA: MIT Press.
Coeckelbergh, M. (2012). *Growing Moral Relations: Critique of Moral Status Ascription*. New York: Palgrave Macmillan.
Cohen, R. A. (2001). *Ethics, Exegesis and Philosophy: Interpretation After Levinas*. Cambridge: Cambridge University Press.
Dawkins, M. S. (2008). The Science of Animal Suffering. *Ethology, 114*(10), 937–945.
Dennett, D. (1998). *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge, MA: MIT Press.
Dennett, D. (1996). *Kinds of Minds: Toward an Understanding of Consciousness*. New York: Basic Books.
Derrida, J. (1978). *Writing and Difference*. Trans. by Alan Bass. Chicago: University of Chicago Press.
Derrida, J. (2005). *Paper Machine. Trans. by Rachel Bowlby*. Stanford, CA: Stanford University Press.
Derrida, J. (2008). *The Animal That Therefore I Am*. Trans. by David Wills. New York: Fordham University Press.
Descartes, R. (1988). *Selected Philosophical Writings. Trans. by J. Cottingham, R. Stoothoff, and D. Murdoch*. Cambridge: Cambridge University Press.
Ess, C. (1996). The Political Computer: Democracy, CMC, and Habermas. In C. Ess (Ed.), *Philosophical Perspectives on Computer-Mediated Communication* (pp. 196–230). Albany, NY: SUNY Press.
Floridi, L. (1999). Information Ethics: On the Philosophical Foundation of Computer Ethics. *Ethics and Information Technology, 1*(1), 37–56.
Floridi, L. (2002). On the Intrinsic Value of Information Objects and the Infosphere. *Ethics and Information Technology, 4*, 287–304.
Floridi, L. (2008). Information Ethics, its Nature and Scope. In J. van den Hoven & J. Weckert (Eds.), *Information Technology and Moral Philosophy* (pp. 40–65). Cambridge: Cambridge University Press.
Floridi, L. (2013). *The Ethics of Information*. Oxford: Oxford University Press.
French, P. (1979). The Corporation as a Moral Person. *American Philosophical Quarterly, 16*(3), 207–215.
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.
Hajdin, M. (1994). *The Boundaries of Moral Discourse*. Chicago: Loyola University Press.
Heidegger, M. (1962). *Being and Time*. Trans. by J. Macquarrie and E. Robinson. New York: Harper and Row Publishers.
Himma, K. E. (2004). There's Something About Mary: The Moral Value of Things qua Information Objects. *Ethics and Information Technology, 6*(3), 145–195.
Himma, K. E. (2009). Artificial Agency, Consciousness, and the Criteria for Moral Agency: What Properties Must an Artificial Agent Have to be a Moral Agent? *Ethics and Information Technology, 11*(1), 19–29.
Kokoro LTD (2009). http://www.kokoro-dreams.co.jp/
Kurzweil, R. (2005). *The Singularity Is Near: When Humans Transcend Biology*. New York: Viking.
Levinas, E. (1969). *Totality and Infinity: An Essay on Exteriority. Trans. by A. Lingis*. Pittsburgh, PA: Duquesne University Press.
Levinas, E. (1981). *Otherwise than Being Or Beyond Essence*. Trans. by Alphonso Lingis. Hague: Martinus Nijhoff Publishers.
Levinas, E. (1987). *Collected Philosophical Papers*. Trans. by A. Lingis. Dordrecht: Martinus Nijhoff Publishers.
Levinas, E. (2003). *Humanism of the Other*. Trans. by Nidra Poller. Urbana: University of Illinois Press.
Llewelyn, J. (1995). *Emmanuel Levinas: The Genealogy of Ethics*. London: Routledge.
Locke, J. (1996). *An Essay Concerning Human Understanding*. Indianapolis, IN: Hackett.
Mauss, M. (1985). *A Category of the Human Mind: The Notion of Person; The Notion of Self. Trans. by W. D. Halls. In M. Carrithers, S. Collins, and S. Lukes (Eds.) The Category of the Person (pp. 1–25)*. Cambridge: Cambridge University Press.
Nealon, J. (1998). *Alterity Politics: Ethics and Performative Subjectivity*. Durham, NC: Duke University Press.
Nietzsche, F. (1974). *The Gay Science*. Trans. by W. Kaufmann. New York: Vintage Books.
Scott, G. E. (1990). *Moral Personhood: An Essay in the Philosophy of Moral Psychology*. Albany, NY: SUNY Press.

Scott, R. L. (1967). On Viewing Rhetoric as Epistemic. *Central States Speech Journal, 18*, 9–17.
Singer, P. (1975). *Animal Liberation: A New Ethics for Our Treatment of Animals*. New York: New York Review of Books.
Singer, P. (1989). All Animals are Equal. In T. Regan & P. Singer (Eds.), *Animal Rights and Human Obligations* (pp. 148–162). New York: Prentice Hall.
Singer, P. (1999). *Practical Ethics*. Cambridge: Cambridge University Press.
Smith, C. (2010). *What Is a Person? Rethinking Humanity, Social Life, and the Moral Good from the Person Up*. Chicago: University of Chicago Press.
Torrence, S. (2008). Ethics and Consciousness in Artificial Agents. *AI & Society, 22*, 495–521.
Velmans, M. (2000). *Understanding Consciousness*. New York: Routledge.
Wallach, W., & Allen, C. (2009). *Moral Machines: Teaching Robots Right from Wrong*. Oxford: Oxford University Press.
Žižek, S. (2006). *The Parallax View*. Cambridge, MA: MIT Press.