Chapter 10

Another Alterity

Rethinking Ethics in the Face of the Machine

David J. Gunkel

In the face of the machine—whether an avatar, chatterbot, or social robot like Nao, Pepper, or Jibo—the question that immediately confronts us is, "Can such artifacts even have face?" Should technology, which has almost always been considered a tool or medium of human action (Heidegger 1977; Feenberg 1991), be considered another socially situated interactive subject? Is it possible for the machine, or even a particular machine, to be Other? In response to these questions, Sherry Turkle identifies what she perceives to be a potentially disturbing trend: "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 2011, 9). In the face of the machine, Turkle argues, we seem to be willing, all too willing, to consider these technological objects to be Other—not just a kind of surrogate pet but a close friend, personal confidant, and even paramour. According to Turkle's diagnosis, we are in danger of substituting the technological interface for the face-to-face encounters we used to have with other human beings. "Technology," she explains, "is seductive when what it offers meets our human vulnerabilities. And as it turns out, we are very vulnerable indeed. We are lonely but fearful of intimacy. Digital connections and the sociable robot may offer the illusion of companionship without the demands of friendship" (Turkle 2011, 1).

As one contemplates the contemporary situation—a situation where, for example, we share intimate details with so-called "friends" on Facebook, spend hours working and reworking the look and appearance of our avatars, or fall for and confide in what turn out to be mindless chatterbots—Turkle's argument definitely appears to be persuasive. But that is, perhaps, the problem. Turkle might have a handle on the individual psychological stressors, but the philosophical opportunities and challenges that appear in the face of this "machinic other" are much more disturbing and interesting. Consequently, the thesis of this chapter can be stated quite simply and directly: The problem with our socially situated and increasingly interactive devices is not that they substitute a machine interface for the face-to-face relationships we used to have with others. Instead, it is in the face of the machine that we are challenged to reexamine critically who or what is, can be, or should be Other. In other words, if the intimacy of network connections and social robots appear to threaten human sociality and communication, it is not simply because these mechanisms are being substituted for real human contact but because it is in facing the question of the alterity of the machine that we are forced to face-up to and reconsider the violent exclusions and defacings that have always been made in the face of others and in the name of ethics.

1. STANDARD OPERATING PRESUMPTIONS

Ethics, in both theory and practice, is an exclusive undertaking. In confronting and dealing with others we inevitably make a decision between "who" is morally significant and "what" remains a mere thing. As Jacques Derrida (2005, 80) explains, the difference between these two small words—"who" and "what"—makes a big difference, precisely because they parse the world into two camps—those Others who count as socially significant and those things which remain mere objects or instruments. These decisions (which are quite literally a cut or "*de-caedere*" in the fabric of being) are often accomplished and justified on the basis of intrinsic, ontological properties. "The standard approach to the justification of moral status is," Mark Coeckelbergh (2012, 13) explains, "to refer to one or more (intrinsic) properties of the entity in question, such as consciousness or the ability to suffer. If the entity has this property, this then warrants giving the entity a certain moral status.” In this transaction, ontology precedes ethics. What something is governs how it is. Or as Luciano Floridi (2013, 116) describes it, “What the entity is determines the degree of moral value it enjoys, if any.” According to this way of thinking—what one might call the standard operating procedure of moral consideration—the question concerning the moral status of others would need to be decided by first identifying which property or properties would be necessary and sufficient to have moral standing and then figuring out whether a particular entity (or class of entities) possesses this property or not. Deciding things in this fashion, although entirely reasonable and expedient, has at least four problems, all of which become increasingly evident and problematic in the face of the machine.

1.1 Substantive Problems

First, how does one ascertain which exact property or properties are necessary and sufficient for moral status? In other words, which one, or ones, count? The history of moral philosophy can, in fact, be read as something of an on-going debate and struggle over this matter with different properties vying for attention at different times. And in this process, many properties that at one time seemed both necessary and sufficient have turned out to be spurious, prejudicial or both. Take for example the following event recalled by Aldo Leopold (1966, 237) at the beginning of “The Land Ethic”: “When god-like Odysseus, returned from the wars in Troy, he hanged all on one rope a dozen slave-girls of his household whom he suspected of misbehavior during his absence. This hanging involved no question of propriety. The girls were property. The disposal of property was then, as now, a matter of expediency, not of right and wrong.” At the time Odysseus is reported to have done this, only male heads of the household were considered legitimate moral and legal subjects. Everything else—his women, his children, and his animals—were property that could be disposed of without any moral hesitation whatsoever. But from where we stand now, the property “male head of the household” is considered a spurious and rather prejudicial criteria for deciding who counts as a moral subject and what remains a mere object.

Similar problems are encounter with, for example, the faculty of reason, which is the property that eventually replaces prejudicial criteria like “male head of the household.” When Immanuel Kant (1985,

17) defined morality as involving the rational determination of the will, non-human animals, which did not possess reason, are categorically excluded from moral consideration. The practical employment of reason does not concern animals and, when Kant does make mention of animality (*Tierheit*), he only uses it as a foil by which to define the boundaries of humanity proper. It is because the human being possesses reason, that he (and the human being, in this particular circumstance, was still principally understood to be male) is raised above the instinctual behavior of the brutes and able to act according to the principles of pure practical reason (Kant 1985, 63).

The property of reason, however, has been subsequently contested by efforts in animal rights philosophy, which begins, according to Peter Singer's analysis, with a critical intervention issued by Jeremy Bentham (2005, 283): "The question is not, 'Can they reason?' nor, 'Can they talk?' but 'Can they suffer?'" According to Singer, the morally relevant property is not speech or reason, which he believes would set the bar for moral inclusion too high, but sentience and the capability to suffer. In *Animal Liberation* (1975) and subsequent writings, Singer argues that any sentient entity, and thus any being that can suffer, has an interest in not suffering and therefore deserves to have that interest taken into account. "If a being suffers," Singer (1975, 9) argues, "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."

This is, however, not the final word on the matter. One of the criticisms of animal rights philosophy, is that this development, for all its promise to intervene in the anthropocentric tradition and include others, still remains an exclusive and exclusionary practice. "If dominant forms of ethical theory," Matthew 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 ethics, for instance, has been critical of animal rights philosophy for organizing its moral innovations on a property (i.e. suffering) that includes some sentient creatures in the community of moral subjects while simultaneously

AQ 14: Is it okay to begin a quoted text with an em dash or should it be changed to ellipses. Please confirm.

justifying the exclusion of other kinds of "lower animals," plants, and the other entities that comprise the natural environment.

But even these efforts to open up and to expand the community of legitimate moral subjects has also (and not surprisingly) been criticized for instituting additional exclusions. "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 artifacts, 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." Consequently, no matter what property (or properties) comes to be identified as morally significant, the choice of property remains contentious, debatable, and seemingly irresolvable. The problem, therefore, is not necessarily deciding which property or properties come to be selected as morally significant. The problem may be in this approach itself, which makes moral consideration dependent upon a prior determination of properties.

1.2 Terminological Troubles

Second, irrespective of which property (or set of properties) is selected, they each have terminological troubles insofar as things like rationality, consciousness, suffering, etc. mean different things to different people and seem to resist univocal definition. Consciousness, for example, is one of the properties that has often been cited as a necessary condition for moral subjectivity (Himma 2009, 19). But consciousness is persistently difficult to define or characterize. The problem, as Max Velmans (2000, 5) points out, is that this term unfortunately "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, ethologists, 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. As Rodney Brooks (2002, 194) admits, "we have no real operational definition of consciousness," and for that reason, "we are completely prescientific at this point about what consciousness is."

Although consciousness, as Anne Foerst remarks, is the secular and supposedly more "scientific" replacement for the occultish "soul" (Benford and Malartre 2007, 162), it appears to be just as much an occult property or what Daniel Dennett (1998, 150) calls an impenetrable "black box."

Other properties do not do much better. Suffering and the experience of pain—which is the property usually deployed in non-standard patient-oriented approaches like animal rights philosophy—is just as problematic, as Dennett cleverly demonstrates in the text, "Why You Cannot Make a Computer That Feels Pain." In this provocatively titled essay, 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" (Dennett 1998, 191). At the end of what turns out to be a rather protracted and detailed consideration of the problem—complete with detailed block diagrams and programming flowcharts—Dennett 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. 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. What Dennett demonstrates, 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 (1998, 228) writes in the conclusion to 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." What Dennett proves, then, is not an inability to program a computer to "feel pain" but our initial and persistent inability to decide and adequately articulate what constitutes "pain" in the first place.

1.3 Epistemological Problems

As if responding to Dennett's challenge, engineers have, in fact, not only constructed mechanisms that synthesize believable emotional responses (Bates 1994; Blumberg, Todd and Maes 1996; Breazeal and Brooks 2004), like the dental-training robot Simroid that cries out in pain when students "hurt" it (Kokoro 2009), but also systems capable of evincing something that appears to be what we generally recognize as “pain.” The interesting issue in these cases is determining whether this is in fact “real pain” or just a simulation. In other words, once the morally significant property or properties have been identified and defined, how can one be entirely certain that a particular entity possesses it, and actually possesses it instead of merely simulating it? Answering this question is difficult, especially because most of the properties that are considered morally relevant tend to be internal mental or subjective states that are not immediately accessible or directly observable. As Paul Churchland (1999, 67) famously asked, “How does one determine 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?” This is, of course, what philosophers commonly call “the problem of other minds,” Though this problem is not necessarily intractable, as I think Steve Torrance (2013) has persuasively argued, the fact of the matter is we cannot, as Donna Haraway (2008, 226) describes it, “climb into the heads of others to get the full story from the inside.”

Although “pain” is not the direct object of his analysis, the epistemological problem of distinguishing between the “real thing” and its mere simulation is illustrated by John Searle’s “Chinese Room.” This influential thought experiment, first introduced in 1980 with the essay “Minds, Brains, and Programs” and elaborated in subsequent publications, was offered as an argument against the claims of strong AI.

> Imagine a native English speaker who knows no Chinese locked in a room full of boxes of Chinese symbols (a data base) together with a book of instructions for manipulating the symbols (the program). Imagine that people outside the room send in other Chinese symbols which, unknown to the person in the room, are questions in Chinese (the input). And imagine that by following the instructions in the program the man in the room is able to pass out Chinese symbols which are correct answers to the questions (the output). The program enables the person in the room to pass the Turing Test for understanding Chinese but he does not understand a word of Chinese. (Searle 1999, 115)

The point of Searle’s imaginative albeit ethnocentric illustration is quite simple—simulation is not the real thing. Merely shifting symbols around in a way that looks like linguistic understanding is not really an understanding of the language. A similar point has been made in the consideration of other properties, like sentience and the experience of pain. Even if, as J. Kevin O'Regan (2007, 332) writes, it were possible to design a robot that "screams and shows avoidance behavior, imitating in all respects what a human would do when in pain …. All this would not guarantee that to the robot, there was actually something it was like to have the pain. The robot might simply be going through the motions of manifesting its pain: perhaps it actually feels nothing at all." The problem exhibited by both examples, however, is not simply that there is a difference between simulation and the real thing. The problem is that we remain persistently unable to distinguish the one from the other in any way that would be considered entirely satisfactory.

1.4 Moral Problems

Finally, there are ethical problems with the properties approach. Any decision concerning qualifying properties is necessarily a normative procedure and an exercise of power over others. In making a determination about the criteria for moral inclusion, someone or some group universalizes their particular experience or situation and imposes this decision on others as the fundamental condition for moral consideration. It is, for example, because human beings experience suffering as both uncomfortable and a moral evil, that it is assumed that the same experience, or at least something substantially similar, in another entity, like an animal, would need to be evaluated and addressed in the same way. This is, for all its promise, still a form anthropocentric thinking and a variety of what the environmental ethicist Thomas Birch (1993, 315) calls "imperial power mongering," insofar as it evaluates the moral standing of others only to the extent that they are "just like us." This is precisely that kind of philosophical thinking that Emmanuel Levinas (1987, 43), criticized for "reducing to the same all that is opposed to it as other."

Consequently "the institution of any practice of any criterion of moral considerability," Birch (1993, 317) argues, "is an act of power over, and ultimately an act of violence toward, those others who turn out to fail the test of the criterion and are therefore not permitted to enjoy the membership benefits of the club of *consideranda.*" In other words, every criteria of moral inclusion, every comprehensive list of qualifying properties, no matter how neutral, objective, or inclusive it appears, is an imposition of power insofar as it consists in the normalization of a particular value or set of values made by someone from a 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." The problem, then, is not only with the specific property or properties that come to be selected as the criteria of moral inclusion but also, and perhaps more so, the very act of selecting properties, which already empowers someone to make these decisions for others.

2. THINKING OTHERWISE

In response to these problems, philosophers—especially in the continental tradition—have advanced alternative approaches to deciding the question of moral standing that can be called, for lack of a better description, "thinking otherwise" (Gunkel 2007). This phrase signifies different ways to formulate the question concerning moral status that is open to and able to accommodate others—and other forms of otherness. And when it comes to thinking otherwise, perhaps no philosopher is better suited to the task than Emmanuel Levinas. Unlike a lot of what goes by the name of "moral philosophy," Levinasian ethics does not get caught up in efforts to determine ontological criteria for inclusion or exclusion but begins from the existential fact that we always and already find ourselves in situations facing and needing to respond to others. For Levinas, ethics transpires not in theorizing about the essential properties of others but in the very real "vulnerabilities," to use Turkle's terminology, that we already experience in the face of others. This change in perspective provides for a number of important innovations that, if pursued far enough, alter how we respond to, and in the face of, others.

2.1 Relatively Relational

According to this alternative way of thinking, moral status is determined and conferred not on the basis of subjective or internal properties decided in advance but according to objectively observable, extrinsic relationships. "Moral consideration," as Coeckelbergh (2010, 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.” As we encounter and interact with other entities—whether they be another human person, an animal, the natural environment, or a domestic robot—this other is first and foremost experienced in relationship to us. The question of moral status, therefore, does not depend on and derive from what the other is in its essence but on how she/he/it (and the choice of pronoun here is part of the problem) stands in relationship to us and how we decide, in the face of the other (to use Levinasian terminology), to respond. Consequently, and contrary to Floridi’s (2013, 116) description, what the entity is *does not* determine the degree of moral value it enjoys. Instead the exposure to the face of the Other, what Levinas calls “ethics,” precedes and takes precedence over all these ontological machinations and determinations. And it is precisely for this reason, that Levinas (1969, 304) famously argued that “morality is not a branch of [applied] philosophy, but first philosophy” where “first” is understood in terms of both temporal sequence and status.

This shift in perspective—a shift that inverts the standard operating procedure by putting ethics before ontology—is not just a theoretical proposal; it has, in fact, been experimentally confirmed in a number of practical investigations with computers and robots. The computer as social actor (CSA) studies undertaken by Byron Reeves and Clifford Nass (1996), for example, demonstrated that human users will accord computers social standing similar to that of another human person and that this occurs as a product of the extrinsic social interaction, irrespective of the actual intrinsic properties (actually known or not) of the entities in question.

AQ 15: Please provide reference details for Byron Reeves and Clifford Nass 1996.

> 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, 22)

In the face of the machine, Reeves and Nass find, human test subjects treat the computer as another socially significant Other. In other words, a significant majority of test subjects respond to the computer as someone "who" counts as opposed to just another "what," and this occurs, they argue, as a product of the extrinsic social circumstances and often in direct opposition to the presumed ontological properties of the mechanism. These results have been verified in two recent studies with robots, one reported in the *International Journal of Social Robotics* (Rosenthal-von der Pütten et al. 2013) where researchers found that human subjects respond emotionally to robots and express empathic concern for machines irrespective of knowledge concerning the properties or inner workings of the device, and another that used physiological evidence, documented by electroencephalography, of the ability of humans to empathize with what appears to be "robot pain" (Suzuki et al. 2015). Although Levinas himself would probably not recognize it as such, what these studies demonstrate is precisely what he had advanced: the ethical response to the other precedes and even trumps decisions concerning ontological properties.

2.2 Radically Superficial

In this situation, the problems of other minds—the difficulty of knowing with any certitude whether the other who confronts us has a conscious mind or is capable of experiencing pain—is not some fundamental epistemological limitation that must be addressed and resolved prior to moral decision making. Levinasian philosophy, instead of being tripped up or derailed by this epistemological problem, immediately affirms and acknowledges it as the condition of possibility for ethics as such. Or, as Richard Cohen succinctly describes it, "not 'other minds,' mind you, but the 'face' of the other, and the faces of all others" (Cohen 2001, 336). 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 wrongheaded) 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."

This means that the order of precedence in moral decision making can and perhaps should be reversed. Internal properties do not come first and then moral respect follows from this ontological fact. Instead the morally significant properties—those ontological criteria that we assume ground moral respect—are what Slavoj Žižek (2008, 209) terms "retroactively (presup)posited" as the result of and as justification for decisions made in the face of social interactions with others. In other words, we project the morally relevant properties onto or into those others who we have already decided to treat as being socially significant—those Others who are deemed to possess face, in Levinasian terminology. In social situations—in contending with the exteriority of others—we always and already decide between "who" counts as morally significant and "what" does not and then retroactively justify these actions by "finding" the properties that we believe motivated this decision making in the first place. Properties, therefore, are not the intrinsic *a prior* condition of possibility for moral standing. They are *a posteriori* products of extrinsic social interactions with and in the face of others.

Once again, this is not some theoretical formulation; it is practically the definition of machine intelligence. Although the phrase "artificial intelligence" is the product of an academic conference organized by John McCarthy at Dartmouth College in 1956, it is Alan Turing's 1950 paper and its "game of imitation," or what is now routinely called "the Turing Test," that defines and characterizes the field. Although Turing begins his essay by proposing to consider the question "Can machines think?" he immediately recognizes the difficulty with defining the subject "machine" and the property "think." For this reason, he proposes to pursue an alternative line of inquiry, one that can, as he describes it, be "expressed in relatively unambiguous words." "The new form of the problem can be described in terms of a game which we call the 'imitation game'. It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman" (Turing 1999, 37). This determination is to be made on the basis of simple questions and answers. The interrogator (C) asks both the man (A) and the woman (B) various questions, and based on their responses (and only their responses) to these inquiries tries to discern whether the respondent is a man or a woman. "In order that tone of voice may not help the interrogator," Turing (1999, 37–38) further stipulates, "the answers should be written, or better still, typewritten.

The ideal arrangement is to have a teleprinter communicating between the two rooms."

Turing then takes his thought experiment one step further. "We can now ask the question, 'What will happen when a machine takes the part of A in this game?' Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, 'Can machines think'?" (Turing 1999, 38). In other words, if the man (A) in the game of imitation is replaced with a computer, would this device be able to respond to questions and simulate the activity of another person? If a computer is capable of successfully simulating a human being in communicative exchanges to such an extent that the interrogator cannot tell whether he is interacting with a machine or another person, then that machine would, Turing concludes, need to be considered "intelligent." Or in Žižek's terms, if the machine effectively passes for another human person in communicative interactions, the property of intelligence would be "retroactively (presup)posited" for that entity, and this is done irrespective of the actual internal states or operations of the other, which are, according to the stipulations of the test, unknown and hidden from view.

2.3 Literally Altruistic

Finally, because ethics transpires in the relationship with others or in the face of the other, decisions about moral standing can no longer be about the granting of rights to others. Instead, the other, first and foremost, questions my rights and challenges my solitude. According to Levinas (1969, 43), "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." This interrupts and even reverses the power relationship enjoyed by previous forms of ethics. Here it is not a privileged group of insiders who then decide to extend rights to others, which is the standard model of all forms of moral inclusion or what Singer (1989, 148) calls a "liberation movement." Instead, the Other challenges and questions the rights and freedoms that I assume I already possess. The principal gesture, therefore, is not the conferring rights on others as a kind of benevolent gesture or even an act of compassion for others but deciding how to respond to the Other, who always and already places my rights and assumed privilege in question. Such an ethics is *altruistic* in the strict sense of the word. It is “of or to others.”

For Levinas, however, this altruism appears to be limited. Whatever the import of his unique contribution, Other in Levinas is still unapologetically human. Although he is not the first to identify it, Jeffrey Nealon 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* [*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, as for many of those who follow in the wake of his influence, the 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 precede ethics.[1]

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 moral innovations. For others, like Matthew Calarco, 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, as I shall argue, 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). 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, 55) argues, “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, 55). This means, of course, that we would be obligated to seriously consider all kinds of others as Other, including other human persons, animals, the natural environment, artifacts, technologies, and robots. An "altruism" that limits in advance who can be Other is not, strictly speaking, altruistic.

3. WHAT OR WHO AND JIBO

We began with the question, "Can the machine have face?" This question, although seemingly direct and intuitive, might not only be the wrong question, but asking it might actually be an impediment to a solution. "There are," as Žižek (2006, 137) explains, "not only true or false solutions, there are also false questions. The task of philosophy is not to provide answers or solutions, but to submit to critical analysis the questions themselves, to make us see how the very way we perceive a problem is an obstacle to its solution." When we ask, "can the machine have face?" we already endorse two problematic assumptions. First, we use the word "machine" as a kind of umbrella term that gathers together all kinds of entities that are not necessarily the same or even similar. Like the word "animal," which gives Derrida (2008) considerable trouble, the word "machine" already makes questionable associations that reduce difference to the same and, at the same time, institutes other potentially problematic distinctions. We could, in fact, repurpose Derrida's exclamation about the word "animal" and apply it to this other word—this word that has been bestowed upon and used to name the other of the animal since at least the time of Descartes: "The machine, what a word! The machine is a word, it is an appellation that men have instituted, a name that they have given themselves the right and authority to give to others" (Derrida 2008, 23). The word "machine," therefore, does not name a neutral ontological category. It is already the product of a crucial decision that has been made in the face of others. In fact, it is precisely by conflating "animal" with "machine" in the hybrid term *bête-machine*, that modern philosophy, beginning with Descartes, succeeded in excluding both animals and machines from moral consideration.[2] Even before one makes a determination between "who" and "what," the machine—by its very name—is already (and along with its other, the animal) located on the side of "what."

Second, the verb "have," as in "have face" or "have a face," has the tendency to turn "face" into a possession and a property that belongs to someone or something. The form of the question, therefore, risks redeploying the properties approach to deciding moral standing in the process of trying to articulate an alternative. Instead of asking "Can the machine *have* face?" we should perhaps rework the form of the question: "What does it take for a machine to supervene and be revealed as Other in the Levinasian sense?" This question—which recognizes, following Coeckelbergh (in this volume), that "alterity is a verb"—no longer asks about "moral standing" in the strict sense of this term, since "standing" suggests that there is an ontological platform onto which morality would be mounted. It therefore comprises a more precise and properly *ethical* question—a question that remains open to others and other forms of otherness: "In what circumstance and under what conditions can a machine—a particular machine that appears here before me—take on face?"

In order to respond to this other question (a question that is otherwise and that can ask about others), we need to consider not "the machine" as a kind of general ontological category but a specific instance of an encounter with a particular entity. In July of 2014 the world got its first look at Jibo. Who or what is Jibo? That is an interesting and important question. In a promotional video that was designed to raise capital investment through pre-orders, social robotics pioneer Cynthia Breazeal introduced Jibo with the following explanation: "This is your car. This is your house. This is your toothbrush. These are your things. But these [and the camera zooms into a family photograph] are the *things* that matter. And somewhere in between is this guy. Introducing Jibo, the world's first family robot" (Jibo 2014). Whether explicitly recognized as such or not, this promotional video leverages Derrida's distinction between "who" and "what." On the side of "what" we have those things that are mere objects—our car, our house, and our toothbrush. According to the instrumental theory of technology, these things are mere instruments that do not have any independent moral status whatsoever (Lyotard 1984, 44). We might worry about the impact that the car's emissions has on the environment (or perhaps stated more precisely, on the health and well-being of the other human beings who share this planet with us), but the car itself is not a moral subject. On the other side there are, as the video describes it "those things that matter." These things are not things, strictly speaking, but are the other persons who count as socially and morally significant Others. Unlike the car, the house, or the toothbrush, these Others have moral status and can be benefited or harmed by our decisions and actions.

Jibo, we are told, occupies a place that is situated somewhere in between *what* are mere things and *who* really matters. Consequently, Jibo is not just another instrument, like our automobile or toothbrush. But he/she/it[3] is also not quite another member of the family pictured in the photograph. Jibo inhabits a place in between these two options. This is, it should be noted, not unprecedented. We are already familiar with other entities who/that occupy a similar ambivalent social position, like the family dog. In fact animals, which since the time of Descartes have been the other of the machine, provide a good precedent for understanding the opportunities and challenges of social robots, like Jibo. Some animals, like the domestic pigs that are raised for food, occupy the position of "what," being mere things that can be used and disposed of as we see fit. Other animals, like a pet dog, are closer to another person "who" counts as Other. They are named, occupy a place alongside us inside the house, and are considered by many to be "a member of the family" (see Gunkel and Coeckelbergh 2014).

As we have seen, we typically theorize and justify the decision between the *what* and the *who* on the basis of intrinsic properties. This approach puts ontology before ethics, whereby what an entity is determines how it comes to be treated. But this method, for all its expediency, also has considerable difficulties: (1) substantive problems with inconsistencies in the identification and selection of the qualifying property, (2) terminological troubles with the definition of the morally significant property, (3) epistemological complications with detecting and evaluating the presence of the property in another, and (4) moral concerns caused by the very effort to use this determination to justify extending moral standing to others. In fact, if we return to the example of animals, it seems very difficult to justify differentiating between the pig, which is a thing we raise and slaughter for food and other raw materials, and the dog, who is a member of the family, on the basis of ontological properties. In terms of all the usual criteria—consciousness, sentience, suffering, etc.—the pig and the dog seem to be virtually indistinguishable. Our moral theory might dictate strict ontological criteria for inclusion and exclusion, but our everyday practices seem to operate otherwise, proving George Orwell's *Animal Farm* (1945, 118) correct: "All animals are equal, but some animals are more equal."

Alternative approaches to making these decisions, like that developed by Levinas, recognize that who is or can be Other is much more complicated. The dog, for instance, occupies the place of an Other who counts, while the pig is excluded as a mere thing, not because of differences in their intrinsic properties, but because of the way these entities have been situated in relationship to us. One of these animals shares our home with us, is bestowed with a proper name, and is considered to have face, to use the Levinasian terminology. The other one does not. Jibo, like an animal, occupies an essentially undecidable position that is in between *who* and *what*. Whether Jibo is or is not an Other, therefore, is not something that will be decided in advance and on the basis of intrinsic properties; it will be negotiated and renegotiated again and again in the face of actual social circumstances. It will, in other words, be out of the actual social relationships we have with Jibo that one will decide whether he/she/it counts or not (and is therefore either a "s/he" or an "it").

Jibo, and other social robots like this, are not science fiction. They are already or will soon be in our lives and in our homes. And in the face of these socially situated and interactive entities, we are going to have to decide whether they are mere things like our car, our house, and our toothbrush; someone who matters like another member of the family; or something altogether different that is situated in between the one and the other. In whatever way this comes to be decided, however, these entities undoubtedly challenge our concept of ethics and the way we typically distinguish between *who* is to be considered Other and *what* is not. Although there are certainly good reasons to be concerned about how these technologies will be integrated into our lives and what the effect of that will be on us, this concern does not justify alarmist reactions and exclusions. We need, following the moral innovations of Levinas, to hold open the possibility that these devices might also implicate us in social relationships where they take on face. At the very least, ethics obligates us—and it does so in advance of knowing anything at all about the inner workings and ontological status of these other kinds of entities—to hold open the possibility that they might become Other. Turkle, therefore, is right about one thing, we are and we should be "willing to seriously consider robots not only as pets but as potential friends, confidants, and even romantic partners." But this is not a dangerous weakness or vulnerability to be avoided at all costs. It is ethics.

NOTES

[1] For Derrida, the anthropocentrism of Levinasian philosophy constitutes cause for considerable concern: “In looking at the gaze of the other, Levinas says, one must forget the color of his eyes, in other words see the gaze, the face that gazes before seeing the visible eyes of the other. But when he reminds us that the ‘best way of meeting the Other is not even to notice the color of his eyes’, he is speaking of man, of one’s fellow as man, kindred, brother; he thinks of the other man and this, for us, will later be revealed as a matter for serious concern” (Derrida 2008, 12). And what truly “concerns” Derrida is not just the way this anthropocentrism limits Levinas’s philosophical innovations but the fact that it already makes exclusive decisions about the (im)possibility of an ethics that is open to and able to accommodate others, like non-human animals.

[2] 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 entities, animals in particular, not only lacked reason but were nothing more than mindless robots that operated on the basis of pre-programmed instructions, like clockwork mechanisms. Conceptualized in this fashion, the animal and the machine were effectively indistinguishable and ontologically the same. (Descartes 1988, 44). Beginning with Descartes, then, the animal and machine share a common form of alterity that situates them as completely different from and distinctly other than human. For more on this association of the animal and the machine, see Derrida (2008) and Gunkel (2012).

[3] The difficult of deciding on a pronoun in this particular situation demonstrates the extent to which efforts to address this other kind of other not only strain against philosophical concepts but also the language available to express such concepts.

REFERENCES

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

Benford, Gregory and Elisabeth Malartre. 2007. *Beyond Human: Living with Robots and Cyborgs*. New York: Tom Doherty.

Bentham, Jeremy. 1780. *An Introduction to the Principles of Morals and Legislation*. Edited by J. H. Burns and H. L. Hart. Oxford: Oxford University Press, 2005.

Birch, Thomas H. 1993. “Moral Considerability and Universal Consideration.” *Environmental Ethics* 15: 313–332.

Birch, Thomas H. 1995. “The Incarnation of Wilderness: Wilderness Areas as Prisons.” In *Postmodern Environmental Ethics.* Edited by Max Oelschlaeger, 137–162. Albany, NY: SUNY Press.

Blumberg, B., P. Todd and M. Maes. 1996. “No Bad Dogs: Ethological Lessons for Learning.” In *Proceedings of the 4th International Conference on Simulation of Adaptive Behavior* (SAB96), 295–304. Cambridge, MA: MIT Press.

Breazeal, Cynthia and Rodney Brooks. 2004. “Robot Emotion: A Functional Perspective.” In *Who Needs Emotions: The Brain Meets the Robot*. Edited by J. M. Fellous and M. Arbib, 271–310. Oxford: Oxford University Press.

Brooks, Rodney A. 2002. *Flesh and Machines*: *How Robots Will Change Us*. New York: Pantheon Books.

Calarco, Matthew. 2008. *Zoographies: The Question of the Animal from Heidegger to Derrida*. New York: Columbia University Press.

Churchland, Paul M. 1999. *Matter and Consciousness*. Cambridge, MIT Press.

Coeckelbergh, Mark. 2010. “Robot Rights? Towards a Social-Relational Justification of Moral Consideration.” *Ethics and Information Technology* 12: 209–221.

Coeckelbergh, Mark. 2012. *Growing Moral Relations*: *Critique of Moral Status Ascription*. New York: Palgrave Macmillan.

Coeckelbergh, Mark and David J. Gunkel. 2014. “Facing Animals: A Relational, Other-Oriented Approach to Moral Standing.” *Journal of Agricultural and Environmental Ethics* 27(5): 715–733.

Cohen, Richard A. 2001. *Ethics, Exegesis, and Philosophy*: *Interpretation After Levinas*. Cambridge: Cambridge University Press.

Dennett, Daniel C. 1998. *Brainstorms*: *Philosophical Essays on Mind and Psychology*. Cambridge, MA: MIT Press.

Derrida, Jacques. 2005. *Paper Machine*. Translated by R. Bowlby. Stanford, CA: Stanford University Press.

Derrida, Jacques. 2008. *The Animal That Therefore I Am*. Edited by Marie-Louise Mallet. Translated by David Wills. New York: Fordham University Press.

Descartes, René. 1988. “Discourse on the Method.” *Selected Philosophical Writings*. Translated by J. Cottingham, R. Stoothoff and D. Murdoch. Cambridge: Cambridge University Press.

Feenberg, Andrew. 1991. *Critical Theory of Technology*. Oxford: Oxford University Press.

Floridi, Luciano. 2013. *The Ethics of Information*. Oxford: Oxford University Press.

Gunkel, David J. 2007. *Thinking Otherwise*: *Philosophy, Communication, Technology*. West Lafayette: Purdue University Press.

Gunkel, David J. 2012. *The Machine Question*: *Critical Perspectives on AI, Robots, and Ethics*. Cambridge, MA: MIT Press.

Haraway, Donna J. 2008. *When Species Meet*. Minneapolis, MN: University of Minnesota Press.

Heidegger, Martin. 1977. “The Question Concerning Technology.” In *The Question Concerning Technology and Other Essays*. Translated by William Lovitt, 3–35. New York: Harper & Row.

Himma, Kenneth Einar. 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.

Jibo. 2014. https://www.jibo.com

Kant, Immanuel. 1985. *Critique of Practical Reason*. Translated by Lewis White Beck. New York: Macmillan.

Kokoro, L. T. D. 2009. http://www.kokoro-dreams.co.jp/

Leopold, Aldo. 1966. *A Sand County Almanac*. Oxford: Oxford University Press.

Levinas, Emmanuel. 1969. *Totality and Infinity.* Translated by Alphoso Lingis. Pittsburgh, PA: Duquesne University Press.

Levinas, Emmanuel. 1987. *Collected Philosophical Papers*. Translated by Alphonso Lingis. Dordrecht: Martinus Nijhoff Publishers.

Lyotard, Jean-François. 1984. *The Postmodern Condition*: *A Report on Knowledge*. Translated by Geoff Bennington and Brian Massumi. Minneapolis, MN: University of Minnesota Press.

Nealon, Jeffrey. 1998. *Alterity Politics*: *Ethics and Performative Subjectivity*. Durham, NC: Duke University Press.

O’Regan, Kevin J. 2007. “How to Build Consciousness into a Robot: The Sensorimotor Approach.” In *50 Years of Artificial Intelligence: Essays Dedicated to the 50th Anniversary of Artificial Intelligence.* Edited by M. Lungarella, F. Iida, J. Bongard and R. Pfeifer, 332–346. Berlin: Springer-Verlag.

Orwell, George. 1945. *Animal Farm*. New York: Harcourt Brace.

Reeves, Byron and Clifford Nass. 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, Astrid M., Nicole C. Krämer, Laura Hoffmann, Sabrina Sobieraj and Sabrina C. Eimler. 2013. “An Experimental Study on Emotional Reactions Towards a Robot.” *International Journal of Social Robotics* 5: 17–34.

Searle, John. 1980. “Minds, Brains, and Programs.” *Behavioral and Brain Sciences* 3(3): 417–457.

Searle, John. 1999. “The Chinese Room.” In *The MIT Encyclopedia of the Cognitive Sciences.* Edited by R. A. Wilson and F. Keil, 115–116. Cambridge, MA: MIT Press.

Singer, Peter. 1975. *Animal Liberation*: *A New Ethics for Our Treatment of Animals*. New York: New York Review of Books.

Singer, Peter. 1989. All animals are equal. In *Animal Rights and Human Obligations*. Edited by Tom Regan and Peter Singer, 148–162. New Jersey: Prentice-Hall.

Suzuki, Yutaka, Lisa Galli, Ayaka Ikeda, Shoji Itakura and Michiteru Kitazaki. 2015. “Measuring Empathy for Human and Robot Hand Pain Using Electroencephalography.” *Scientific Reports* 5: 15924. http://www.nature.com/articles/srep15924

Torrance, Steve. 2013. “Artificial Consciousness and Artificial Ethics: Between Realism and Social Relationism.” *Philosophy & Technology* 27(1): 9–29.

Turing, Alan M. 1999. “Computing Machinery and Intelligence.” In *Computer Media and Communication.* Edited by P. A. Mayer, 37–58. Oxford: Oxford University Press.

Turkle, Sherry. 2011. *Alone Together: Why We Expect More from Technology and Less From Each Other*. New York: Basic Books.

Velmans, Max. 2000. *Understanding Consciousness*. London, UK: Routledge.

Žižek, Slavoj. 2006. “Philosophy, The ‘Unknown Knowns’, And the Public Use of Reason.” *Topoi* 25: 137–142.

Žižek, Slavoj. 2008. *For They Know Not What They Do: Enjoyment as a Political Factor*. London: Verso.