The Rights of Machines—Caring for Robotic Care-Givers

David J. Gunkel[1]

Abstract. Work in the field of machine medical ethics, especially as it applies to healthcare robots, generally focuses attention on controlling the decision making capabilities and actions of autonomous machines for the sake of respecting the rights of human beings. Absent from much of the current literature is a consideration of the other side of this issue. That is, the question of machine rights or the moral standing of these socially situated and interactive technologies. This chapter investigates the moral situation of healthcare robots by examining how human beings should respond to these artificial entities that will increasingly come to care for us. A range of possible responses will be considered bounded by two opposing positions. We can, on the one hand, deal with these mechanisms by deploying the standard instrumental theory of technology, which renders care-giving robots nothing more than tools and therefore something we do not really need to care about. Or we can, on the other hand, treat these machines as domestic companions that occupy the place of another "person" in social relationships, becoming someone we increasingly need to care about. Unfortunately neither option is entirely satisfactory, and it is the objective of this chapter not to argue for one or the other but to formulate the opportunities and challenges of ethics in the era of robotic caregivers.

1 INTRODUCTION

Work in the field of machine medical ethics, especially as it applies to increasingly autonomous home healthcare robots, generally focuses attention on the capabilities, modes of implementation, and range of actions of these mechanisms for the sake of respecting the dignity and rights of human patients. Researchers like Noel and Amanda Sharkey, for instance, have focused on a spectrum of potentially troubling moral problems: infantalization of those under robotic care; deception, especially with regards to individuals suffering from impaired or diminished mental/emotional capabilities [1]; and "the rights to privacy, personal liberty and social contact" [2]. Others, like Robert and Linda Sparrow, question whether increasing involvement of robots in elder care (an area of research that the Sparrows argue is second only to military applications) would in fact be adequate to meet not just the looming demographic crisis of an aging population but the complex emotional and social needs of seniors [3]. And Mark Coeckelbergh has focused on the concept of care itself, asking whether machines can be adequately designed and implemented to supply what, under normal circumstances, would constitute not just acceptable but "good care" [4]. In all these cases what is at issue is the well-being and rights of human patients and the extent to which machines improve or adversely affect human flourishing. "The primary concern," as Borenstein and Pearson describe it, "is about how the existence of robots may positively or negatively affect the lives of care recipients" [5].

Absent from much of the current literature, however, is a consideration of the other side of this issue, namely the moral status and standing of these machines. Unlike the seemingly cold and rather impersonal industrial robots that have been successfully developed for and implemented in manufacturing, transportation, and maintenance operations, home healthcare robots will occupy a unique social position and "share physical and emotional spaces with the user" [6]. In providing care for us, these machines will take up residence in the home and will be involved in daily personal and perhaps even intimate interactions (i.e. monitoring, feeding, bathing, mobility, and companionship). For this reason, it is reasonable to inquire about the social status and moral standing of these technologies. How, for example, will human patients under the care of such mechanisms respond to these other entities? How should we respond to them? What are or what will be our responsibilities to these others—another kind of socially aware and interactive other? The following takes up and investigates this "machine question" by examining the moral standing of robots, and home healthcare robots in particular. Because there are a number of different and competing methods by which to formulate and decide this question, the chapter will not supply one definitive answer, but will consider a number of related moral perspectives that, taken together, add up to an affirmative response to the question concerning the rights of machines.

2 DEFAULT SETTING

From a traditional philosophical perspective, the question of machine rights or machine moral standing not only would be answered in the negative but the query itself risks incoherence. "To many people," David Levy writes, "the notion of robots having rights is unthinkable" [7]. This is because machines are assumed to be nothing more than instruments of human activity and have no independent moral status whatsoever. This common sense determination is structured and informed by the answer that, as Martin Heidegger argues, is typically provided for the question concerning technology:

> We ask the question concerning technology when we ask what it is. Everyone knows the two statements that answer our question. One says: Technology is a means to an end. The other says: Technology is a human activity. The two definitions of technology belong together. For to posit ends and procure and utilize the means to them is a human activity. The manufacture and utilization of equipment, tools, and machines, the manufactured and used things themselves, and the needs and ends that they serve, all belong to what technology is [8].

[1] Department of Communication, Northern Illinois University, DeKalb, IL 60115, USA. Email: dgunkel@niu.edu / http://gunkelweb.com

According to Heidegger's analysis, the presumed role and function of any kind of technology, whether it be the product of handicraft or industrialized manufacture, is that it is a means employed by human users for specific ends. Heidegger termed this particular characterization "the instrumental and anthropological definition" and indicated that it forms what is considered to be the "correct" understanding of any kind of technological contrivance [8].

Under this clearly human-centered formulation, technology, no matter how sophisticated its design or operations, is considered to be nothing more than a tool or instrument of human endeavor. As Deborah Johnson explains, "computer systems are produced, distributed, and used by people engaged in social practices and meaningful pursuits. This is as true of current computer systems as it will be of future computer systems. No matter how independently, automatic, and interactive computer systems of the future behave, they will be the products (direct or indirect) of human behavior, human social institutions, and human decision" [9]. Understood in this way, machines—even those machines that are programmed to care for us—are not legitimate moral subjects that we need to care about. They are neither moral agents responsible for actions undertaken by or through their instrumentality nor moral patients, that is, the recipients of action and the subject of moral considerability. "We have never," as J. Storrs Hall points out, "considered ourselves to have moral duties to our machines, or them to us" [10].

On this account, the bar for machine moral standing appears to be impossibly high if not insurmountable. In order for a machine to have anything like "rights," it would need to be recognized as human or at least virtually indistinguishable from another human being in social situations and interactions. Although this has often been the subject of science fiction—consider, for example, Isaac Asimov's short story "The Bicentennial Man," in which the android "Andy" seeks to be recognized as legally human—it is not limited to fictional speculation, and researchers like Hans Moravec [11], Rodney Brooks [12], and Raymond Kurzweil [13] predict human-level or better machine capabilities by the middle of the century. Although achievement of this remains hypothetical, the issue is not necessarily whether machines will or will not attain human-like capabilities. The problem resides in the anthropocentric criteria itself, which not only marginalizes machines but has often been instrumental for excluding other human beings. "Human history," Christopher Stone argues, "is the history of exclusion and power. Humans have defined numerous groups as less than human: slaves, woman, the 'other races,' children and foreigners. These are the wretched who have been defined as stateless, personless, as suspect, as rightsless" [14].

Because of this, recent innovations have sought to disengage moral standing from this anthropocentric privilege and have instead referred matters to the generic concept "person." "Many philosophers," Adam Kadlac argues, "have contended that there is an important difference between the concept of a person and the concept of a human being" [15]. One such philosopher is Peter Singer. "Person," Singer writes, "is often used as if it meant the same as 'human being.' Yet the terms are not equivalent; there could be a person who is not a member of our species. There could also be members of our species who are not persons" [16]. In 2013, for example, India declared dolphins "non-human persons, whose rights to life and liberty must be respected" [17]. Likewise corporations 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 [18]. And not surprisingly, there has been, in recent years, a number of efforts to extend the concept "person" to AI's, intelligent machines, and robots [19, 20, 21].

As promising as this innovation appears to be, however, there is little agreement concerning what makes someone or something a person, and the literature on this subject is littered with different formulations and often incompatible criteria [15, 16, 22, 23]. In an effort to contend with, if not resolve these problems, researchers often focus on the one "person making" quality that appears on most, if not all, the lists—consciousness. "Without consciousness," John Locke famously argued, "there is no person" [24]. For this reason, consciousness is widely considered to be a necessary if not sufficient condition for moral standing, and there has been considerable effort in the fields of philosophy, AI, and robotics to address the question of machine moral standing by targeting the possibility (or impossibility) of machine consciousness [25, 26].

This determination is dependent not only on the design and performance of actual artifacts but also—and perhaps more so—on how we understand and operationalize the term "consciousness." Unfortunately there has been little or no agreement concerning this matter, and the concept encounters both terminological and epistemological problems. First, we do not have any widely accepted definition of "consciousness," and the concept, as Max Velmans points out "means many different things to many different people" [27]. In fact, if there is any agreement among philosophers, psychologists, cognitive scientists, neurobiologists, AI researchers, and robotics engineers regarding this matter, it is that there is little or no agreement, when it comes to defining and characterizing the term. To make matters worse, the difficulty 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 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" [28].

Second, even if it were possible to define consciousness or come to some agreement (no matter how tentative or incomplete) as to what characterizes it, 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 inaccessible. "How does one determine," as Paul Churchland 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?" [29]. Although philosophers, psychologists, and neuroscientists throw considerable argumentative and experimental effort at this problem—and much of it is rather impressive and persuasive—it is not able to be fully and entirely resolved. Consequently, not only are we

unable to demonstrate with any certitude whether animals, machines, or other entities are in fact conscious (or not) and therefore legitimate moral persons (or not), we are left with doubting whether we can, without fudging the account, even say the same for other human beings. And it is this persistent and irreducible difficulty that opens the space for entertaining the possibility of extending rights to other entities like machines or animals.

3 BÊTE-MACHNE

The situation of animals is, in this context, particularly interesting and important. Animals have not traditionally been considered moral subjects, and it is only recently that the discipline of philosophy has begun to approach the animal as a legitimate target of moral concern. The crucial turning point in this matter is derived from a brief but influential statement provided by Jeremy Bentham: "The question is not, Can they reason? nor Can they talk? but, Can they suffer?" [30]. Following this insight, the crucial issue for animal rights philosophy is not to determine whether some entity can achieve human-level capacities with things like speech, reason, or consciousness; "the first and decisive question would be rather to know whether animals can suffer" [31].

This change in perspective—from a standard agent-oriented to a non-standard patient-oriented ethics [32]–provides a potent model for entertaining the question of the moral standing and rights of machines. This is because the animal and the machine, beginning with the work of René Descartes, share a common ontological status and position—marked, quite literally in the Cartesian texts, by the hybrid term *bête-machine* [33]. Despite this essential similitude, animal rights philosophers have resisted efforts to extend rights to machines, and they demonize Descartes for even suggesting the association [34]. This exclusivity has been asserted and justified on the grounds that the machine, unlike an animal, is not capable of experiencing either pleasure or pain. Like a stone or other inanimate object, a mechanism would have nothing that mattered to it and therefore, unlike a mouse or other sentient creature would not be a legitimate subject of moral concern, because "nothing that we can do to it could possibly make any difference to its welfare" [35]. Although this argument sounds rather reasonable and intuitive, it fails for at least three reasons.

First, it has been practically disputed by the construction of various mechanisms that now appear to exhibit emotional responses or at least provide external evidence of behaviors that effectively simulate and look undeniably like pleasure or pain. As Derrida 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'" [31]. 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 rights philosophy, 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 [36, 37, 38, 39], like the dental-training robot Simroid "who" cries out in pain when students "hurt" it [40], and designed systems capable of evidencing behaviors that look a lot like what we usually call pleasure and pain. Although programming industrial robots with emotions—or, perhaps more precisely stated, the capability to simulate emotions—would be both unnecessary and perhaps even misguided, this is something that would be desirable for home healthcare robots, which will need to exhibit forms of empathy and emotion in order to better interact with patients and support their care.

Second, it can be contested on epistemologically grounds. Because suffering is typically understood to be subjective, there is no way to know exactly how another entity experiences unpleasant (or pleasant) sensations. Like "consciousness," suffering is also an internal state of mind and is therefore complicated by the problem of other minds. As Singer 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" [35]. The basic problem, then, is not whether the question "Can they suffer?" also applies to machines but whether anything that appears to suffer—human, animal, plant, or machine—actually does so at all. Furthermore, and to make matters even more complex, 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 in Daniel Dennett's, "Why You Can't Make a Computer That Feels Pain." In this provocatively titled essay, 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" [41]. 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. The reason you cannot make a computer that feels pain, Dennett argues, 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.

Third, all this talk about the possibility of engineering pain or suffering in order to demonstrate machine rights entails its own particular moral dilemma. "If (ro)bots might one day be capable of experiencing pain and other affective states," Wendell Wallach and Colin Allen 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? [42] If it were in fact possible to construct a robot that "feels pain" (however defined and instantiated) in order to demonstrate the moral standing of machines, 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. If it were in fact possible to construct a device that "feels pain" in order to demonstrate the possibility of machine moral standing, then doing so might be ethically

problematic insofar as in constructing such a mechanism we do not do everything in our power to minimize its suffering. Or to put it another way, positive demonstration of "machine rights," following the moral innovations and model of animal rights philosophy, might only be possible by risking the violation of those rights.

4 THINKING OTHERWISE

Irrespective of how it is articulated, these different approaches to deciding moral standing focus on what Mark Coeckelbergh calls "(intrinsic) properties" [43]. This method is rather straight forward and intuitive: "identify one or more morally relevant properties and then find out if the entity in question has them" [43]. But as we have discovered, there are at least two persistent problems with this undertaking. First, how does one ascertain which exact property or properties are 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—rationality, speech, consciousness, sentience, suffering, etc.—vying for attention at different times. Second, once the morally significant property (or properties) has been identified, how can one be entirely certain that a particular entity possesses it, and actually possesses it instead of merely simulating it? This is tricky business, 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. In response to these problems, there are two alternatives that endeavor to consider and address things otherwise.

4.1 Machine Ethics

The first concerns what is now called Machine Ethics. This relatively new idea was first introduced and publicized in a 2004 AAAI paper written by Michael Anderson, Susan Leigh Anderson, and Chris Armen and has been followed by a number of dedicated symposia and publications [44, 45]. Unlike computer ethics, which is interested in the consequences of human behaviour through the instrumentality of technology, "machine ethics is concerned," as characterized by Anderson et al. "with the consequences of behaviour of machines toward human users and other machines" [46]. In this way, machine ethics both challenges the "human-centric" tradition that has persisted in moral philosophy and argues for a widening of the subject so as to take into account not only human action with machines but also the behaviour some machines, namely those that are designed to provide advice or programmed to make autonomous decisions with little or no human supervision. And for the Andersons, healthcare applications provide both a test case and occasion for the development of working prototypes.

Toward this end, machine ethics takes an entirely functionalist approach to things. That is, it considers the effect of machine actions on human subjects irrespective of metaphysical debates concerning moral standing or epistemological problems concerning subjective mind states. As Susan Leigh Anderson points out, the Machine Ethics project is unique insofar as it, "unlike creating an autonomous ethical machine, will not require that we make a judgment about the ethical status of the machine itself, a judgment that will be particularly difficult to make" [47]. Machine Ethics, therefore, does not necessarily deny or affirm the possibility of, for instance, machine personhood, consciousness, or sentience. It simply endeavors to institute a pragmatic approach that does not require that one first decide these questions a priori. It leaves these matters as an open question and proceed to ask whether moral decision making is computable and whether machines can in fact be programmed with appropriate ethical standards for acceptable forms of social behavior.

This is a promising innovation insofar as it recognizes that machines are already making decisions and taking real-world actions in such a way that has an effect—and one that can be evaluated as either good or bad—on human beings and human social institutions. Despite this, the functionalist approach utilized by Machine Ethics has at least three critical difficulties. First, functionalism shifts attention from the cause of an action to its effects.

> Clearly relying on machine intelligence to effect change in the world without some restraint can be dangerous. Until fairly recently, the ethical impact of a machine's actions has either been negligible, as in the case of a calculator, or, when considerable, has only been taken under the supervision of a human operator, as in the case of automobile assembly via robotic mechanisms. As we increasingly rely upon machine intelligence with reduced human supervision, we will need to be able to count on a certain level of ethical behavior from them [46].

The functionalist approach instituted by Machine Ethics derives from and is ultimately motivated by an interest to protect human beings from potentially hazardous machine decision-making and action. This effort, despite arguments to the contrary, is thoroughly and unapologetically anthropocentric. Although effectively opening up the community of moral subjects to other, previously excluded things, the functionalist approach only does so in an effort to protect human interests and investments. This means that the project of Machine Ethics does not differ significantly from computer ethics and its predominantly instrumental and anthropological orientation. If computer ethics, as Anderson et al. characterize it, is about the responsible and irresponsible use of computerized tools by human users [46], then their functionalist approach is little more than the responsible design and programming of machines by human beings for the sake of protecting other human beings.

Second, functionalism institutes, as the conceptual flipside and consequence of this anthropocentric privilege, what is arguably a slave ethic. "I follow," Kari Gwen Coleman writes, "the traditional assumption in computer ethics that computers are merely tools, and intentionally and explicitly assume that the end of computational agents is to serve humans in the pursuit and achievement of their (i.e. human) ends. In contrast to James Gips' call for an ethic of equals, then, the virtue theory that I suggest here is very consciously a slave ethic" [48]. For Coleman, computers and other forms of computational agents should, in the words of Joanna Bryson, "be slaves" [49]. Others, however, are not so confident about the prospects and consequences of this "Slavery 2.0." Concern over this matter is something that is clearly exhibited and developed in robot science fiction from *R.U.R.* and *Metropolis* to *Bladerunner* and *Battlestar Galactica.* But it has also been expressed by

contemporary researchers and engineers. Rodney Brooks, for example, recognizes that there are machines that are and will continue to be used and deployed by human users as instruments, tools, and even servants. But he also recognizes that this approach will not cover all machines in all circumstances.

> Fortunately we are not doomed to create a race of slaves that is unethical to have as slaves. Our refrigerators work twenty-four hours a day seven days a week, and we do not feel the slightest moral concern for them. We will make many robots that are equally unemotional, unconscious, and unempathetic. We will use them as slaves just as we use our dishwashers, vacuum cleaners, and automobiles today. But those that we make more intelligent, that we give emotions to, and that we empathize with, will be a problem. We had better be careful just what we build, because we might end up liking them, and then we will be morally responsible for their well-being [50].

According to Brooks's analysis, a slave ethic will work, and will do so without any significant moral difficulties or ethical friction, as long as we decide to produce dumb instruments that serve human users as mere tools or extensions of our will. But as soon as the machines show signs, however minimal defined or rudimentary, that we take to be intelligent, conscious, or intentional, then everything changes. What matters here, it is important to note, is not the actually capabilities of the machines but the way we read, interpret, and respond to their actions and behaviours. As soon as we see what we think are signs of something like intelligence, intentionality, or emotion, a slave ethic will no longer be functional or justifiable.

Finally, even those seemingly unintelligent and emotionless machines that can legitimately be utilized as "slaves" pose a significant ethical problem. This is because machines that are designed to follow rules and operate within the boundaries of some kind of programmed restraint, might turn out to be something other than a neutral tool. Terry Winograd, for example, warns against something he calls "the bureaucracy of mind," "where rules can be followed without interpretive judgments" [51]. Providing robots, computers, and other autonomous machines with functional morality produces little more than artificial bureaucrats—decision making mechanisms that can follow rules and protocols but have no sense of what they do or understanding of how their decisions might affects others. "When a person," Winograd argues, "views his or her job as the correct application of a set of rules (whether human-invoked or computer-based), there is a loss of personal responsibility or commitment. The 'I just follow the rules' of the bureaucratic clerk has its direct analog in 'That's what the knowledge base says.' The individual is not committed to appropriate results, but to faithful application of procedures" [51]. Coeckelbergh paints an even more disturbing picture. For him, the problem is not the advent of "artificial bureaucrats" but "psychopathic robots" [4]. The term "psychopathy" refers to a kind of personality disorder characterized by an abnormal lack of empathy which is masked by an ability to appear normal in most social situations. Functional morality, Coeckelbergh argues, intentionally designs and produces what are arguably "artificial psychopaths"—robots that have no capacity for empathy but which follow rules and in doing so can appear to behave in morally appropriate ways. These psychopathic machines would "follow rules but act without fear, compassion, care, and love. This lack of emotion would render them non-moral agents—i.e. agents that follow rules without being moved by moral concerns—and they would even lack the capacity to discern what is of value. They would be morally blind" [4].

4.2 Social Relational Ethics

An alternative to moral functionalism can be found in Coeckelbergh's own work, where he develops an approach to moral status ascription that he characterizes as "social relational" [43]. By this, he means to emphasize the way moral status is not something located in the inner recesses or essential make-up of an individual entity but transpires through actually existing interactions and relationships situated between entities. This "relational turn," which Coeckelbergh develops by capitalizing on innovations in ecophilosophy, Marxism, and the work of Bruno Latour, Tim Ingold, and others, does not get bogged down trying to resolve the philosophical problems associated with the standard properties approach. Instead it recognizes the way that moral status is socially constructed and operationalized. Quoting the environmental ethicist Baird Callicot, Coeckelbergh insists that the "relations are prior to the things related" [43]. This almost Levinasian gesture is crucial insofar as it reconfigures the usual way of thinking. It is an anti-Cartesian and post-modern (in the best sense of the word) intervention. In Cartesian modernism the individual subject had to be certain of his (and at this time the subject was always gendered male) own being and essential properties prior to engaging with others. Coeckelbergh reverses this standard approach. He argues that it is the social that comes first and that the individual subject (an identity construction that is literally thrown under or behind this phenomena), only coalesces out of the relationship and the assignments of rights and responsibilities that it makes possible.

This relational turn in moral thinking is clearly a game changer. As we interact with machines, whether they be pleasant customer service systems, medical advisors, or home healthcare robots, the mechanism is first and foremost situated and encountered in relationship to us. Morality, conceived of in this fashion, is not determined by a prior ontological determination concerning the essential capabilities, intrinsic properties, or internal operations of these other entities. Instead it is determined in and by the way these entities come to face us in social interactions. Consequently, "moral consideration is," as Coeckelbergh describes it, "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" [4]. This is the reason why, as Levinas claims, "morality is first philosophy" ("first" in terms of both sequence and status) and that moral decision making precedes ontological knowledge [52]. Ethics, conceived of in this way, is about decision and not discovery [53]. We, individually and in collaboration with each other (and not just those others who we assume are substantially like ourselves), decide who is and who is not part of the moral community—who, in effect, will have been admitted to and included in this first person plural pronoun.

This is, it is important to point out, not just a theoretical proposal but has been experimentally confirmed in a number of empirical investigations. The computer as social actor (CSA) studies undertaken by Byron Reeves and Clifford Nass and reported in their influential book *The Media Equation*,

demonstrate that human users will accord computers social standing similar to that of another human person. This occurs, as Reeves and Nass demonstrate, as a product of the social interaction and irrespective of the actual ontological properties (actually known or not) of the machine in question [54]. Similar results have been obtained by Christopher Bartneck et al and reported in the paper "Daisy, Daisy, Give me your answer do! Switching off a Robot," a title which refers to the shutting down of the HAL 9000 computer in Stanley Kubrick's *2001: A Space Odyssey*. In Bartneck et al's study, human subjects interacted with a robot on a prescribed task and then, at the end of the session, were asked to switch off the machine and wipe its memory. The robot, which was in terms of its programming no more sophisticated than a basic chatter bot, responded to this request by begging for mercy and pleading with the human user not to shut it down. As a result of this, Bartneck and company recorded considerable hesitation on the part of the human subjects to comply with the shutdown request [55]. Even though the robot was "just a machine"—and not even very intelligent—the social situation in which it worked with and responded to human users, made human beings consider the right of the machine (or at least hesitate in considering this) to continued existence.

For all its opportunities, however, this approach to deciding moral standing otherwise is inevitably and unavoidably exposed to the charge of moral relativism—"the claim that no universally valid beliefs or values exist" [56]. To put it rather bluntly, if moral status is relational and open to different decisions concerning others made at different times for different reasons, are we not at risk of affirming an extreme form of moral relativism? One should perhaps answer this indictment not by seeking some definitive and universally accepted response (which would obviously reply to the charge of relativism by taking refuge in and validating its opposite), but by following Slavoj Žižek's strategy of "fully endorsing what one is accused of" [57]. So yes, relativism, but an extreme and carefully articulated form of it. That is, a relativism that can no longer be comprehended by that kind of understanding of the term which makes it the mere negative and counterpoint of an already privileged universalism. Relativism, therefore, does not necessarily need to be construed negatively and decried, as Žižek himself has often done, as the epitome of postmodern multiculturalism run amok [58]. It can be understood otherwise. "Relativism," Robert Scott argues, "supposedly, means a standardless society, or at least a maze of differing standards, and thus a cacophony of disparate, and likely selfish, interests. Rather than a standardless society, which is the same as saying no society at all, relativism indicates circumstances in which standards have to be established cooperatively and renewed repeatedly" [59]. In fully endorsing this form of relativism and following through on it to the end, what one gets is not necessarily what might have been expected, namely a situation where anything goes and "everything is permitted" [60]. Instead, what is obtained is a kind of ethical thinking that turns out to be much more responsive and responsible in the face of others.

5 CONCLUSION

In November of 2012, General Electric launched a television advertisement called "Robots on the Move." The 60 second spot, created by Jonathan Dayton and Valerie Faris (the husband/wife team behind the 2006 feature film *Little Miss Sunshine*), depicts many of the iconic robots of science fiction traveling across great distances to assemble before some brightly lit airplane hanger for what we are told is the unveiling of some new kind of machines—"brilliant machines," as GE's tagline describes it. And as we observe Robby the Robot from *Forbidden Planet*, KITT the robotic automobile from *Knight Rider*, and Lt. Commander Data of *Star Trek: The Next Generation* making their way to this meeting of artificial minds, we are told, by an ominous voice over, that "the machines are on the move."

Although this might not look like your typical robot apocalypse (vividly illustrated in science fiction films and television programs like *Terminator*, *The Matrix Trilogy*, and *Battlestar Galactica*), we are, in fact, in the midst of an invasion. The machines are, in fact, on the move. They may have begun by displacing workers on the factory floor, but they now actively participate in many aspects of social life and will soon be invading and occupying places in our homes. This invasion is not some future possibility coming from a distant alien world. It is here. It is now. And resistance appears to be futile. What matters for us, therefore, is how we decide to respond to this opportunity/challenge. And in this regard, we will need to ask some important but rather difficult questions: At what point might a robot, or other autonomous system be held fully accountable for the decisions it makes or the actions it deploys? When, in other words, would it make sense to say "It's the robot's fault"? Likewise, at what point might we have to consider seriously extending rights—civil, moral, and legal standing—to these socially aware and interactive devices that will increasingly come to serve and care for us, our children, and our aging parents? When, in other words, would it no longer be considered non-sense to suggest something like "the rights of robots"?

In response to these questions, there appears to be at least two options, neither of which are entirely comfortable or satisfactory. On the one hand, we can respond as we typically have, treating these mechanisms as mere instruments or tools. Bryson makes a reasonable case for this approach in her essay "Robots Should be Slaves": "My thesis is that robots should be built, marketed and considered legally as slaves, not companion peers" [49]. Although this moral imperative (marked, like all imperatives, by the verb "should") might sound harsh, this line of argument is persuasive, precisely because it draws on and is underwritten by the instrumental theory of technology—a theory that has considerable history and success behind it and that functions as the assumed default position for any and all considerations of technology. This decision—and it is a decision, even if it is the default setting—has both advantages and disadvantages. On the positive side, it reaffirms human exceptionalism in ethics, making it absolutely clear that it is only the human being who possess rights and responsibilities. Technologies, no matter how sophisticated, intelligent, and influential, are and will continue to be mere tools of human action, nothing more. But this approach, for all its usefulness, has a not-so-pleasant downside. It willfully and deliberately produces a new class of instrumental servants or slaves and rationalizes this decision as morally appropriate and justified. In other words, applying the instrumental theory to these new kinds of domestic healthcare machines, although seemingly reasonable and useful, might have devastating consequences for us and others.

On the other hand, we can decide to entertain the possibility of rights and responsibilities for machines just as we had previously done for other non-human entities, like animals [35], corporations [18], and the environment [61]. And there is both moral and legal precedent for this transaction. Once again, this decision sounds reasonable and justified. It extends moral standing to these other socially active entities and recognizes, following the predictions of Norbert Wiener, that the social situation of the future will involve not just human-to-human interactions but relationships between humans and machines [62]. But this decision also has significant costs. It requires that we rethink everything we thought we knew about ourselves, technology, and ethics. It requires that we learn to think beyond human exceptionalism, technological instrumentalism, and all the other -isms that have helped us make sense of our world and our place in it. In effect, it calls for a thorough reconceptual-ization of who or what should be considered a legitimate moral subject and risks involvement in what is often considered antithetical to clear moral thinking—relativism.

In any event, how we respond to the opportunities and challenges of this machine question will have a profound effect on the way we conceptualize our place in the world, who we decide to include in the community of moral subjects, and what we exclude from such consideration and why. But no matter how it is decided, it is a decision—quite literally a cut that institutes difference and makes a difference. And, as Blay Whitby correctly points out, the time to start thinking about and debating these issues is now...if not yesterday [63].

REFERENCES

[1] Sharkey, Noel and Amanda Sharkey. Granny and the robots: ethical issues in robot care for the elderly. *Ethics and information technology* 14(1): 27-40 (2012).

[2] Sharkey, Noel and Amanda Sharkey. The rights and wrongs of robot care. In *Robot Ethics: The Ethical and Social Implications of Robotics*, ed. Patrick Lin, Keith Abney and George A. Bekey, 267-282. Cambridge, MA: MIT Press (2012).

[3] Sparrow, Robert and Linda Sparrow. In the hands of machines? The future of aged care. *Minds and Machines* 16: 141–161 (2010).

[4] Coeckelbergh, Mark. Moral appearances: emotions, robots, and human morality. *Ethics and Information Technology* 12(3): 235-241 (2010).

[5] Borenstein, Jason and Yvette Pearson. Robot caregivers: ethical issues across the human lifespan. In *Robot Ethics: The Ethical and Social Implications of Robotics*, ed. Patrick Lin, Keith Abney and George A. Bekey, 251-265. Cambridge, MA: MIT Press (2012).

[6] Cerqui, D. and K. O. Arras. Human beings and robots: towards a symbiosis? In *A 2000 People Survey. Post-Conference Proceedings PISTA 03*. (Politics and Information Systems: Technologies and Applications), ed. J. Carrasquero, 408-413 (2001)

[7] Levy, David. *Robots Unlimited: Life in Virtual Age.* Wellesley, MA: A K Peters (2008).

[8] Heidegger, Martin. *The Question Concerning Technology and Other Essays.* Trans. William Lovitt. New York: Harper & Row (1977).

[9] Johnson, Deborah G. Computer systems: moral entities but not moral agents. *Ethics and Information Technology* 8: 195-204 (2006).

[10] Hall, J. Storrs. Ethics for machines. In *Machine Ethics*, ed Michael Anderson and Susan Leigh Anderson, 28-44. Cambridge: Cambridge University Press (2011).

[11] Moravec, Hans. *Mind Children: The Future of Robot and Human Intelligence*. Cambridge, MA: Harvard University Press (1988).

[12] Brooks, Rodney. *Flesh and Machines: How Robots Will Change Us*. New York: Pantheon (2002).

[13] Kurzweil, Raymond. *The Singularity is Near: When Humans Transcend Biology*. New York: Viking (2005).

[14] Stone, Christopher D Should trees have standing? toward legal rights for natural objects. *Southern California Law Review* 44: 450-492 (1972).

[15] Kadlac, Adam. Humanizing personhood. *Ethical Theory and Moral Practice* 13(4): 421-437 (2009).

[16] Singer, Peter. *Practical Ethics.* Cambridge: Cambridge University Press (1999).

[17] Coelho,Saroja. Dolphins gain unprecedented protection in india. *Deutsche Welle*. http://dw.de/p/18dQV (2013).

[18] French, Peter. The corporation as a moral person. *American Philosophical Quarterly* 16(3): 207-215 (1979).

[19] Hubbard, F. Patrick. Do androids dream?: Personhood and intelligent artifacts. *Temple Law Review* 83: 101-170 (2011).

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

[21] Peterson, Steve. Designing people to serve. In *Robot Ethics: The Ethical and Social Implications of Robotics*, ed. Patrick Lin, Keith Abney and George A. Bekey, 282-298. Cambridge, MA: MIT Press (2012).

[22] Ikäheimo, Heikki and Arto Laitinen. Dimensions of personhood: editors' introduction. *Journal of Consciousness Studies* 14(5-6): 6-16 (2007).

[23] Smith, Christian. *What is a Person? Rethinking Humanity, Social Life, and the Moral Good from the Person up.* Chicago: University of Chicago Press (2010).

[24] Locke, John. *An Essay Concerning Human Understanding*. Indianapolis, IN: Hackett (1996).

[25] Himma, Kenneth Einar. 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 (2009).

[26] Torrance, Steve. Ethics and consciousness in artificial agents. *AI & Society* 22: 495-521 (2008).

[27] Velmans, Max. *Understanding Consciousness.* London, UK: Routledge (2000).

[28] Güzeldere, Güven. The many faces of consciousness: a field guide. In *The Nature of Consciousness: Philosophical Debates*, ed. Ned Block, Owen Flanagan and Güven Güzeldere, 1-68. Cambridge, MA: MIT Press (1997).

[29] Churchland, Paul M. *Matter and Consciousness* (revised edition). Cambridge, MA: MIT Press (1999).

[30] Bentham, Jeremy. *An Introduction to the Principles of Morals and Legislation*, edited by J H Burns and H L Hart. Oxford: Oxford University Press (2005).

[31] Derrida, Jacques. *The Animal That I Therefore Am*. Trans. David Wills. New York: Fordham University Press (2008).

[32] Floridi, Luciano and J. W. Sanders. On the morality of artificial agents. *Minds and Machine* 14: 349-379 (2004).

[33] Descartes, Rene. *Selected Philosophical Writings*. Trans. John Cottingham, Robert Stoothoff and Dugald Murdoch. Cambridge: Cambridge University Press (1988).

[34] Regan, Tom. *The Case for Animal Rights*. Berkeley, CA: University of California Press (1983).

[35] Singer, Peter. *Animal Liberation: A New Ethics for our Treatment of Animals.* New York: New York Review Book (1975).

[36] Bates, J. The role of emotion in believable agents. *Communications of the ACM* 37: 122-125 (1994).

[37] Blumberg, B, P. Todd and M. Maes. No bad dogs: ethological lessons for learning. *Proceedings of the 4th International Conference on Simulation of Adaptive Behavior* (SAB96). Cambridge, MA: MIT Press, p 295-304 (1996).

[38] Breazeal, Cynthia and Rodney Brooks. Robot emotion: a functional perspective. In *Who Needs Emotions: The Brain Meets the Robot*, ed. J. M. Fellous and M. Arbib, 271-310. Oxford: Oxford University Press (2004).

[39] Velásquez, Juan D. When robots weep: emotional memories and decision-making. *AAAI-98 Proceedings*. Menlo Park, CA: AAAI Press (1988).
[40] Kokoro LTD. http://www.kokoro-dreams.co.jp/ (2009).
[41] Dennett, Daniel C. *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge, MA: MIT Press (1988).
[42] Wallach, Wendell and Colin Allen. *Moral Machines: Teaching Robots Right from Wrong*. Oxford: Oxford University Press (2009).
[43] Coeckelbergh, Mark. *Growing Moral Relations: Critique of Moral Status Ascription.* New York: Palgrave MacMillan (2012).
[44] Anderson, Michael and Susan Leigh Anderson Machine ethics. *IEEE Intelligent Systems* 21(4): 10-11 (2006).
[45] Anderson, Michael and Susan Leigh Anderson. *Machine Ethics*. Cambridge: Cambridge University Press. (2011).
[46] Anderson, Michael, Susan Leigh Anderson and Chris Armen. Toward machine ethics. *American Association for Artificial Intelligence.* http://www.aaai.org/Papers/ Workshops/2004/WS-04-02/WS04-02-008.pdf (2004).
[47] Anderson, Susan Leigh. Asimov's "three laws of robotics" and machine metaethics. *AI & Society* 22(4): 477-493 (2008).
[48] Coleman, Kari Gwen. Android arete: toward a virtue ethic for computational agents." Ethics and Information Technology 3: 247-265 (2001).
[49] Bryson, Joanna. Robots Should be Slaves. In *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*, ed. Yorick Wilks, 63-74. Amsterdam: John Benjamins (2010).
[50] Brooks, Rodney. *Flesh and Machines: How Robots Will Change Us*. New York: Pantheon (2002).
[51] Winnograd, Terry. Thinking machines: can there be? are we? In *The Foundations of Artificial Intelligence: A Sourcebook*. ed. Derek Partridge and Yorick Wilks, 167-189. Cambridge: Cambridge University Press (1990).
[52] Levinas, Emmanuel. *Totality and Infinity: An Essay on Exteriority*. Trans. Alphonso Lingis. Pittsburgh, PA: Duquesne University Press (1969).
[53] Putnam, Hilary. Robots: machines or artificially created life? *The Journal of Philosophy* 61(21): 668-691 (1964).
[54] Reeves, Byron and Clifford Nass. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places.* Cambridge: Cambridge University Press (1996).
[55] Bartneck, Christoph, Michel van der Hoek, Omar Mubin and Abdullah Al Mahmud. Daisy, daisy, give me your answer do! Switching off a robot. *Proceedings of the 2nd ACM/IEEE International Conference on Human-Robot Interaction*. Washington, DC, 217-222 (2007).
[56] Ess, Charles. The political computer: Democracy, CMC, and Habermas. In *Philosophical Perspectives on Computer-Mediated Communication*, ed. Charles Ess, 197-231. Albany: State University of New York Press (1996).
[57] Žižek, Slavoj. *The Fragile Absolute or, Why is the Christian Legacy Worth Fighting For?* New York: Verso (2000).
[58] Žižek, Slavoj. *The Puppet and the Dwarf: The Perverse Core of Christianity.* Cambridge, MA: MIT Press (2003).
[59] Scott, Robert L. On viewing rhetoric as epistemic. *Central States Speech Journal* 18: 9-17 (1967).
[60] Camus, Albert. *The Myth of Sisyphus, and Other Essays.* Trans. J O'Brien. New York: Alfred A. Knopf (1983).
[61] Birch, Thomas. Moral considerability and universal consideration. *Environmental Ethics* 15: 313-332 (1993).
[62] Wiener, Norbert. *The Human use of Human Beings.* New York: Da Capo (1954).
[63] Whitbey, Blay. Sometimes it's hard to be a robot: a call for action on the ethics of abusing artificial agents. *Interacting with Computers* 20(3): 326-333 (2008).