Thinking otherwise: Ethics, technology and other subjects

David J. Gunkel
Department of Communication, Northern Illinois University, DeKalb, IL, 60115, USA
E-mail: dgunkel@niu.edu;

Abstract. Ethics is ordinarily understood as being concerned with questions of responsibility for and in the face of an other. This other is more often than not conceived of as another human being and, as such, necessarily excludes others – most notably animals and machines. This essay examines the ethics of such exclusivity. It is divided into three parts. The first part investigates the exclusive anthropocentrism of traditional forms of moral thinking and, following the example of recent innovations in animal rights philosophy, questions the mechanisms of such exclusion. Although recent work in animal- and bio-ethics has successfully implemented strategies for the inclusion of the animal as a legitimate subject of moral consideration, its other, the machine, has remained conspicuously excluded. The second part looks at recent attempts to include these machinic others in moral thinking and critiques the assumptions, values, and strategies that have been employed by these various innovations. And the third part proposes a means for thinking otherwise. That is, it introduces an alternative way to consider these other forms of otherness that is not simply reducible to the conceptual order that has structured and limited moral philosophy's own concern with and for others.

Key words: computer ethics, computers-social aspects, Emmanuel Levinas, otherness, philosophy of technology, machine ethics

Introduction

> In addition, we might ask about those ethical calls of the future from "beings" that we cannot now even imagine.[1] – Jeffrey T. Nealon

This essay concerns a question. The question has to do with ethics. And, like so much contemporary work on this subject matter, it is interested in the call from, our response to, and our responsibility for others. However, unlike the capital "O" other, who has taken center stage in recent moral thinking largely due to the influence of Emmanuel Levinas and others, this investigation is concerned with and directed otherwise. It is interested in those others who remain, for reasons that will need to be explicated, excluded from the contemporary and essentially anthropocentric understandings of alterity that have persisted in and been constitutive of moral philosophy. It is, therefore, a question that is not concerned with or limited to the present and the presence of these somewhat familiar Others. It is a question that is oriented to the future, toward other "beings," who we may not now even be able to imagine and who call to us and approach from elsewhere. It is, in short, a question

about the future of ethics. And it involves and addresses itself to the possibility of an other who remains fundamentally and disturbingly otherwise.

An example, might help to illustrate this rather abstract characterization. In a now well-known and often reproduced *New Yorker* cartoon by Peter Steiner, two dogs sit in front of an Internet-connected personal computer. The one operating the machine says to his companion, "On the Internet, nobody knows you're a dog."[2] The cartoon has often been cited to address issues of identity and anonymity in computer-mediated communication.[3] As Richard Holeton interprets it, "the cartoon makes fun of the anonymity of network

---

[1] Jeffrey T. Nealon, *Alterity Politics: Ethics and Performative Subjectivity* (Durham, NC: Duke University Press, 1998), 71.

[2] Peter Steiner, "Dog cartoon," *The New Yorker* (5 July 1993): 61.

[3] See for instance, Roger F. Fidler, *Mediamorphosis: Understanding New Media* (Thousand Oaks, CA: Pine Forge Press, 1997); Richard Holeton, ed., *Composing Cyberspace: Identity, Community, and Knowledge in the Electronic Age* (New York: McGraw Hill, 1998); Sara Kiesler, ed., *Culture of the Internet* (Mahwah, NJ: Lawrence Erlbaum Associates, 1997); William J. Mitchell, *City of Bits: Space, Place, and the Infobahn* (Cambridge: MIT Press, 1995); Geoffrey Nunberg, "Prefixed Out," commentary on *Fresh Air*, WHYY radio, 17 May 2002. Transcript available at http://www-csli.stanford.edu/~nunberg/cyber.html; Diana Saco, *Cybering Democracy: Public Space and the Internet* (Minneapolis, MN: University of Minnesota Press, 2002).

communications by showing a dog online, presumably fooling some credulous humans about its true identity."[4] According to this particular reading, what the cartoon portrays is that who or what one *is* in computer-mediated communication (CMC) is, as Sandy Stone, Sherry Turkle, and others have demonstrated, something that can be easily and endlessly reconfigured.[5] This interpretation of the cartoon, although not necessarily incorrect, misses the more interesting and suggestive insight that is provided by the wired canines. What the cartoon demonstrates is not the anonymity and indeterminacy of others in CMC but the unquestioned assumption that despite this anonymity, users assume that the other with whom they interact is another human. The other who confronts us in cyberspace is always, it is assumed, another human being, like ourselves. These others may be "other" in a "celebrate diversity" sense of the word – another race, another gender, another ethnicity, another social class, etc. But they are never a dog. Consequently, what the cartoon shows, through a kind of clever inversion, is the standard operating presumption of mainstream moral philosophy and much of computer ethics. Online identity is, in fact, reconfigurable. You can be a dog, or you can say you are. But everyone knows, or so it is assumed, that what is on the other end of the fiber-optic cable is another human user, someone who is, despite what are often interpreted as minor variations in physical appearance and background, essentially like we assume ourselves to be. The cartoon works, because in CMC everyone already assumes that the other is human. "Inside the little box," Stone concludes, "are other people."[6]

The following investigates and seeks to intervene in this deep-seated and often unquestioned assumption, tracing its limits and their effect on our understanding and application of ethics. The investigation is divided into three sections. The first examines the anthropocentric bias of traditional forms of moral philosophy and the moral repercussions of excluding others – specifically the animal and the machine. The second considers and critiques three recent innovations – machine ethics, android ethics, and information ethics – which purport to expand the purview of moral philosophy but do so by adhering to and redeploying similar, if not the same, exclusionary practices. The third and final section responds to this fundamental limitation by proposing an alternative approach that is addressed and organized *otherwise*.

---

[4] Holeton, 111.

[5] Allucquère Rosanna Stone, *The War of Desire and Technology at the Close of the Mechanical Age* (Cambridge, MA: MIT Press, 1995); Sherry Turkle, *Life on the Screen: Identity in the Age of the Internet* (New York: Simon & Schuster, 1995).

[6] Stone, 16.

Exclusion

> Humanism administers lessons to 'us' (?). In a million ways, often mutually incompatible. Well founded (Apel) and non-founded (Rorty), counterfactual (Habermas, Rawls) and pragmatic (Searl), psychological (Davidson) and ethico-political (the French neo-humanists). But always as if at least man were a certain value, which has not need to be interrogated.[7] – Jean-François Lyotard

Ethics is customarily understood as being concerned with questions of responsibility for and in the face of an other. For traditional forms of moral philosophy, this "other" is more often than not conceived of as another human being – another human subject who is essentially and necessarily like we assume ourselves to be. This anthropocentrism necessarily excludes others, most notably the animal and the machine. In fact, it is through the systemic exclusion of these others that the human as human has come to be defined, delimited, and characterized. Although this kind of exclusivity is enacted and described throughout the history of western thought, it is perhaps most evident in the work of René Descartes. For Descartes, the human being is the sole creature capable of rational thought. In this view, animals not only lack reason but are nothing more than mindless automatons that, like a clockwork mechanism, follow predetermined instructions that are programmed in the disposition of their various parts or organs. Understood in this way, the animal and machine become virtually indistinguishable. "If any such machine," Descartes writes, "had the organs and outward shape of a monkey or of some other animal that lacks reason, we should have no means of knowing that they did not possess entirely the same nature as these animals."[8] Consequently, the animal and machine share a common form of alterity that situates them as completely different from and distinctly other than human.

Because of this exclusion from the realm of rational thought, the animal has not traditionally been considered a legitimate moral subject. When Kant, for example, defined morality as involving the rational

---

[7] Jean-François Lyotard, *The Inhuman: Reflections on Time*, trans. Geoffrey Bennington and Rachel Bowlby (Stanford, CA: Stanford University Press, 1991), 1.

[8] René Descartes, *Discourse on Method*, in *Descartes: Selected Philosophical Writings*, trans. and eds. John Cottingham, Robert Stoothoff and Dugald Murdoch (Cambridge: Cambridge University Press, 1988), 44. In 1738 this characterization was practically demonstrated, when Jacques de Vaucanson exhibited a mechanical duck, which reportedly was indistinguishable from a real duck.

determination of the will,[9] the animal, which does not by definition possess reason, is immediately and categorically excluded. The practical employment of reason does not concern the animal and, when Kant does make mention of animality [*Tierheit*], he only uses it as a foil by which to define the limits of humanity proper.[10] It is because the human being possesses reason, that he (and the human being, in this case, was principally male) is raised above the brute instinctual behavior of mere animality and able to act according to the principles of pure practical reason.[11] The same ethical redlining is effected in the analytic tradition. According to Tom Regan, this is immediately apparent in the seminal work of analytical ethics. "It was in 1903 when analytic philosophy's patron saint, George Edward Moore, published his classic, *Principia Ethica*. You can read every word in it. You can read between every line of it. Look where you will, you will not find the slightest hint of attention to 'the animal question.' Natural and nonnatural properties, yes. Definitions and analyses, yes. The open-question argument and the method of isolation, yes. But so much as a word about non-human animals? No. Serious moral philosophy, of the analytic variety, back then did not traffic with such ideas."[12] This exclusive anthropocentrism is also at work in the philosophy of Emmanuel Levinas,[13] the most notable ethicist in the continental tradition. Unlike a lot of what goes by the name of "moral philosophy," Levinasian ethics does not rely on metaphysical generalizations, abstract formulas, or simple pieties. His philosophy is concerned with the response to and responsibility for the absolutely Other who is confronted in an irreducible face-to-face encounter. Whatever the import of this unique contribution, this other is always and unapologetically human. Although he is not the first to identify this problem, Jeffrey Nealon provides what is perhaps the most succinct description of it in *Alterity Politics*: "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* [ἄνθρωποX] and only *anthropos*, has *logos* [λόγοX]; 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*."[14] For Levinas, as for those modes of ethical thinking that follow in the wake of his influence, the other is always operationalized as another human subject.[15] If, as Levinas argues, ethics precedes ontology, then in Levinas' own work anthropology and a certain brand of humanism precede ethics.

It is only recently that the discipline of philosophy has begun to approach the animal as a legitimate subject of ethics. Regan identifies the turning point in a single work: "In 1971, three Oxford philosophers –

---

[9] Immanuel Kant, *Critique of Practical Reason*, trans. Lewis W. Beck (New York: Macmillan, 1985), 17.

[10] According to this reading, Kantian philosophy merely excludes the animal from consideration. Theodore Adorno, as Derrida points out in the final essay of *Paper Machine* (Stanford, CA: Stanford University Press, 2005), takes the interpretation one step further, arguing that Kant not only excluded animality but held it in contempt. "He [Adorno] particularly blames Kant, whom he respects too much from another point of view, for not giving any place in his concept of dignity (*Würde*) and the 'autonomy' of man to any compassion (*Mitleid*) between man and the animal. Nothing is more odious (*verhasster*) to Kantian man, says Adorno, than remembering a resemblance or affinity between man and animal (*die Erinnerung an die Tierähnlichkeit des Menschen*). The Kantian feels only hate for human animality" (180).

[11] Kant, 63.

[12] Tom Regan, forward to *Animal Others: On Ethics, Ontology, and Animal Life*, ed. Peter Steeves (Albany, NY: State University of New York Press, 1999), xii.

[13] Emmanuel Levinas, *Totality and Infinity*, trans. Alphonso Lingis (Pittsburgh, PA: Duquesne University Press, 1969); Emmanuel Levinas, *Otherwise Than Being or Beyond Essence*, trans. Alphonso Lingis (The Hague: Martinus Nijhoff Publishers, 1981).

[14] Nealon, 71.

[15] Levinas's humanism is also identified and addressed by Derrida in the introduction to his 1997 presentation at Cerisy-la-Salle and in Richard Cohen's introduction to the English translation of Levinas's 1972 publication *Humanism of the Other*. For Derrida ("The Animal That Therefore I Am," *Critical Inquiry* 28, winter 2002), the humanist pretensions of Levinasian philosophy constitute 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" (381). Whereas Derrida maintains a critical stance toward Levinas's humanism, Cohen's introduction to *Humanism of the Other* (Chicago: University of Illinois Press, 2003) gives it a positive spin: "The three chapters of *Humanism of the Other* each defend humanism – the world view founded on the belief in the irreducible dignity of humans, a belief in the efficacy and worth of human freedom and hence also of human responsibility" (ix). This is not the place to engage in this debate concerning Levinasian philosophy. However, what is important to note for the analysis at hand is the fact that both sides of the argument recognize and affirm a fundamental humanism always and already at work within Levinas's ethics of otherness.

Roslind and Stanley Godlovitch, and John Harris – published *Animals, Men and Morals*. The volume marked the first time philosophers had collaborated to craft a book that dealt with the moral status of nonhuman animals."[16] According to Regan, this particular publication is not only credited with introducing what is now called the "animal question," but launched an entire sub-discipline of moral philosophy where the animal is considered to be a legitimate subject of ethical inquiry. Currently, philosophers of both the analytic and continental varieties[17] find reason to be concerned with animals, and there is a growing body of research addressing issues like the ethical treatment of animals, animal rights, and environmental ethics. According to Cary Wolfe, there are two factors that have made this remarkable reversal of the tradition possible. On the one hand, there is the crisis of humanism, "brought on, in no small part, first by structuralism and then post-structuralism and its interrogation of the figure of the human as the constitutive (rather than technically, materially, and discursively constituted) stuff of history and the social."[18] Since at least Nietzsche, philosophers, anthropologists, and social scientists have been increasingly suspicious of the privileged position human beings have given themselves in the great chain of being, and this suspicion becomes an explicit object of inquiry within the so called "human sciences." On the other hand, the boundary between the animal and the human has, as Donna Haraway remarks, become increasingly untenable. Everything that had divided us from them is now up for grabs: language, tool use, and even reason.[19] Recent discoveries in various branches of the biological sciences have had the effect of slowly dismantling the wall that

Descartes and others had erected between the human and the animal. According to Wolfe, "a veritable explosion of work in areas such as cognitive ethology and field ecology has called into question our ability to use the old saws of anthropocentrism (language, tool use, the inheritance of cultural behaviors, and so on) to separate ourselves once and for all from the animals, as experiments in language and cognition with great apes and marine mammals, and field studies of extremely complex social and cultural behaviors in wild animals such as apes, wolves, and elephants, have more or less permanently eroded the tidy divisions between human and nonhuman."[20] The revolutionary effect of this transformation can be seen, somewhat ironically, in the backlash of what Evan Ratliff calls "creationism 2.0," a well organized "crusade against evolution" that attempts to reinstate a clear and undisputed division between human beings and the rest of terrestrial life based on a strict interpretation of the Judeo-Christian creation myth.[21]

What is curious is that at a time when this other form of otherness is increasingly recognized as a legitimate subject of moral philosophy, its other, the machine, remains conspicuously absent. Despite all the talk of the animal question, animal others, animal rights, and the reconsideration of what Wolfe calls the "repressed Other of the subject, identity, logos,"[22] virtually nothing is said about the machine. One could, in fact, redeploy Regan's critique of G. E. Moore's *Principia Ethica* and apply it, with a high degree of accuracy, to any work purporting to address the animal question: "You can read every word in it. You can read between every line of it. Look where you will, you will not find the slightest hint of attention to 'the machine question.'" Even though the fate of the machine, from Descartes on, was intimately coupled with that of the animal, only one of the pair has qualified for ethical consideration. This exclusion is not just curious; it is illogical and indefensible. In fact, it seems as if the machine, even before the animal, should have challenged the anthropocentric prejudice that is the operating system of western ethics. Unlike the animal, the machine, especially the information processing machine that comprises so much of contemporary technology, appears to possess something like intelligence, reason, or *logos*. Not only can these machines engage in the complexities of mathematics, which for Descartes, Leibniz, Kant, and others constituted the

---

[16] Regan, xi.

[17] In continental thought, the prominence of "the animal question" is evident in Martin Heidegger's 1929–1930 lecture course *Die Grundbegriffe der Metaphysik: Welt – Endlichkeit – Einsamkeit* (Frankfurt am Main: V. Klostermann, 1983); David Farrell Krell's commentary on this text in *Daimon Life: Heidegger and Life Philosophy* (Bloomington, IN: Indiana University Press, 1992); Giorgio Agamben's *The Open: Man and Animal* (Stanford, CA: Stanford University Press, 2004); and the lectures delivered at the third Cerisy-la-Salle conference (1997) and subsequently published under the title *L'Animal autobiographique, Autour de Jacques Derrida* (Paris: Editions Galilée, 1999).

[18] Cary Wolfe, introduction to *Zoontologies: The Question of the Animal*, ed. Cary Wolfe (Minneapolis, MN: University of Minnesota Press, 2003), x–xi.

[19] Donna Haraway, *Simian, Cyborgs and Women: The Reinvention of Nature* (New York: Routledge, 1991), 151–152.

[20] Wolfe, xi.

[21] Evan Ratliff, "The Crusade Against Evolution," *Wired* 12.10 (October 2004): 156–161.

[22] Wolfe, x.

epitome of rational thought and the proper model of philosophy, but simple chatter-bots like Joseph Weizenbaum's ELIZA can apparently participate in intelligent dialogue, arranging words in such a way as to provide logical and meaningful responses.[23] Despite this, it is only the animal that has qualified for ethical consideration. Despite all the ink spilled on the subject of the animal question, almost nothing (and this important qualification will be examined shortly) has been written about the machine. And despite all the talk about an ethics of radical otherness, we have said little or nothing about a machinic other. "We have," in the words of J. Storrs Hall, "never considered ourselves to have 'moral' duties to our machines, or them to us."[24] How can we continue to justify this exclusion? If we admit animals, do we not also have to admit the machine? Can an ethics that is oriented toward the other get away with including one and not the other? Can such an ethics persist without being exposed as inconsistent, capricious, and, in a word, unethical? The choice is clear, but each option seems difficult and problematic. Either we own up to the exclusive strategy of ethics, continue to redline the machine, and install new mechanisms to dispel the hypocrisy that will inevitably threaten such a maneuver at every turn. Or we open the flood gates and admit that it now makes sense, perhaps had always made sense, to entertain the machine question and consider the moral rights of machines. Either way, ethics will never be the same.

Inclusion

Would an ethics be sufficient, as Levinas maintains, to remind the subject of its being-subject, its being-guest, host or hostage, that is to say its being-subjected-to-the-other, to the Wholly Other or to every single other? I don't think so. It takes more than that to break with the Cartesian tradition of the animal-machine that exists without language and without the ability to respond.[25] – Jacques Derrida

Despite what appears to be systemic exclusion, a small number of scholars have begun to give serious attention to the machine and its position, or lack thereof, in moral philosophy. Although these innovations do not provide definitive answers to the machine question, they do demonstrate the kind of thinking and discursive maneuvers that are necessary to respond to and to take responsibility for the machine as a legitimate moral subject. At the same time, however, these efforts, insofar as they are indebted to the language and logic of the tradition, also and unavoidably reinscribe traditional values and assumptions and, by doing so, demonstrate by their own practices just how difficult it is to think and articulate an ethics that is oriented otherwise. This section will consider and critique three recent innovations: machine ethics, android ethics, and information ethics. "Critique," however, is a word that is not without ambiguity. In colloquial usage, it often has a negative connotation, indicating a form of judgmental evaluation or rudimentary fault-finding. There is, however, a more precise definition that is rooted in the tradition of critical philosophy. "A critique of any theoretical system," as Barbara Johnson characterizes it, "is not an examination of its flaws and imperfections. It is not a set of criticisms designed to make the system better. It is an analysis that focuses on the grounds of that system's possibility."[26] The following comprises this kind of operation. As such, the analysis of machine, android, and information ethics does not simply identify and examine the flaws and imperfections of these alternative theories. Instead, it focuses on their condition of possibility, demonstrating that these alternatives, although purporting to open up ethics to other possibilities, remain grounded in the same problematic assumptions and exclusionary practices.

One attempt to think moral philosophy beyond its traditionally limited configuration can be found in what is now called "machine ethics." This relatively new idea was first introduced and publicized in a paper written by Michael Anderson, Susan Leigh Anderson, and Chris Armen and presented during the 2004 Workshop on Agent Organizations held in conjunction with the American Association for

---

[23] Not only are there machines that are capable of conversing with human users in a way that is virtually indistinguishable from another person, but the machine has even been determined to outpace the human in this distinctly human occupation. In the estimation of Kevin Warwick ("Cyborg Morals, Cyborg Values, Cyborg Ethics," *Ethics and Information Technology* 5.3, 2003), "the biggest advantage of all for machine intelligence is communication. In comparison with the capabilities of machines, human communication is so poor as to be embarrassing" (132).

[24] J. Storrs Hall, "Ethics for Machines," *KurzweilAI.net* (5 July 2001), http://www.kurzweilai.net/articles/art0218.html

[25] Jacques Derrida, "And Say the Animal Responded," trans. David Willis, in *Zoontologies: The Question of the Animal*, ed. Cary Wolfe (Minneapolis, MN: University of Minnesota Press, 2003), 121.

[26] Barbara Johnson, "Translator's Introduction" in Jacques Derrida, *Disseminations* (Chicago: University of Chicago Press, 1981), xv.

Artificial Intelligence's (AAAI) nineteenth national conference. This debut, which appropriately sought "to lay the theoretical foundation for *machine ethics*"[27] was followed with the formation of the Machine Ethics Consortium (MachineEthics.org) and a dedicated 2005 AAAI symposium on the subject.[28] Unlike computer ethics, which is mainly concerned with the consequences of human behavior through the instrumentality of computer technology, "*machine ethics* is concerned," as characterized by Anderson et al., "with the consequences of behavior of machines toward human users and other machines."[29] 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 of ethics so as to take into account not only human action with machines but the behavior of some machines, namely those that are designed to provide advice or programmed to make autonomous decisions with little or no human supervision. "Clearly," Anderson and company write, "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."[30]

Although significantly expanding the subject of ethics by incorporating the subjectivity and agency of machines, machine ethics does not, it is important to note, provide any consideration of our response to and responsibility for these machinic others. In other words, machine ethics is exclusively interested in articulating ethical guidelines and procedures for the way machines deal with and treat human beings. This approach, according to Anderson et al., is motivated by an interest to protect human beings from potentially hazardous machine decisions and actions. Deploying machine intelligence in the world without some kind of preprogrammed ethical restraint is, on their account, potentially dangerous for the human species. Consequently, the project of machine ethics, like Isaac Asimov's three laws of robotics,[31] is motivated by a desire to manage the potential hazards of intelligent machines for the sake of ensuring the humane treatment of human beings. At the theoretical foundation of machine ethics, then, is an affirmation of the supreme and unquestioned value of the human. This affirmation is precisely the organizing assumption of anthropocentric ethics, which machine ethics had purported to put in question and to suspend. Despite its critique of the anthropocentrism that dominates moral philosophy in general and computer ethics in particular, machine ethics deploys and reinforces a human-centered perspective. As a result, it is not situated outside of and as an alternative to computer ethics but remains part and parcel of that tradition. If computer ethics is, as Anderson et al. characterize it, about the responsible and irresponsible use of the computer by human users, then machine ethics is little more than the responsible programming of machines by human beings for the sake of protecting other human beings. Instead of laying the foundation for a new moral perspective, machine ethics redeploys the anthropocentric prejudice through the mechanism of what initially appears to be a

---

[27] Michael Anderson, Susan Leigh Anderson, and Chris Armen, "Toward Machine Ethics," American Association for Artificial Intelligence – The Nineteenth National Conference on Artificial Intelligence, 25–29 July 2004, San Jose, CA, 1. It should be emphasized that by "debut" I mean the institution of "machine ethics" as a recognized and distinct area of study within the discipline of artificial intelligence (AI). The motivations and ideas behind machine ethics had already been available and articulated within the discipline of AI for quite some time prior to this. See for example, the 2000 Convention of the Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB-00) Symposium on Artificial Intelligence, Ethics and (Quasi-) Human Rights (University of Birmingham, UK, 19–20 April 2000), especially Blay Whitby and Kane Oliver's "How to Avoid a Robot Takeover: Political and Ethical Choices in the Design and Introduction of Intelligent Artifacts," Helen Seville and Debora Field's "What Can AI Do for Ethics?" and Joanna Bryson's "A Proposal for the Humanoid Agent-builders League (HAL)."

[28] For a summary and evaluation of work presented at the AAAI 2005 Fall Symposium on Machine Ethics, see Michael Anderson and Susan Leigh Anderson, "The Status of Machine Ethics: A Report from the AAAI Symposium," *Mind and Machines*, forthcoming. Preprint available at http://www.springerlink.com/content/e502572456857542

[29] Anderson et al., 1. Whether this clarification is adequate to distinguish "computer ethics" from what Anderson et al. call "machine ethics" is something that is questionable and remains open to considerable debate.

[30] Ibid., 4.

[31] Asimov, *I, Robot* (New York: Bantam Books, 1991). Recently Susan Leigh Anderson ("Asimov's 'Three Laws of Robotics' and Machine Metaethics," *AI and Society*, forthcoming. Preprint available at http://www.springerlink.com/content/771k1181268772p1) has endeavored to distinguish the project of machine ethics from Asimov's three laws of robotics, arguing that the later is unable to provide a satisfactory basis for the former.

critique and alternative. This is not necessarily some deliberate deception instituted by the authors or even an accidental lapse in thinking. It is, on the contrary, the result of and evidence for the almost complete saturation that has been achieved by the humanist perspective in ethical matters. In other words, the anthropocentrism that has characterized the last 2500+ years of moral philosophy is so pervasive and inescapable that any attempt to think outside the humanist box, like that of Anderson and his colleagues, is already and unavoidably caught in the language, logic, and protocols of this legacy system.

Although machine ethics is concerned exclusively with the moral consequences of actions performed by machines, other theorists have considered whether and to what extent machines are deserving of ethical consideration as such. Robert Sparrow, for instance, foresees the need for something he calls "Android Ethics." "As soon as AIs begin to possess consciousness, desires and projects then it seems as though they deserve some sort of moral standing."[32] In order to define the ethical tipping point – the point at which a computer becomes the appropriate subject of moral concern – Sparrow proposes a modification of the Turing Test. The modified test, like the original, is proposed as a thought experiment, which "asks when a computer might fill the role of a human being in a moral dilemma."[33] The dilemma selected by Sparrow is the case of triage.

> In the scenario I propose, a hospital administrator is faced with the decision as to which of two patients on life support systems to continue to provide electricity to, following a catastrophic loss of power in the hospital. She can only preserve the existence of one and there are no other lives riding on the decision. We will know that machines have achieved moral standing comparable to a human when the replacement of one of the patients with an artificial intelligence leave the character of the dilemma intact. That is, when we might sometimes judge that it is reasonable to preserve the continued existence of the machine over the life of the human being. This is the "*Turing Triage Test*."[34]

As it is described by Sparrow, the "Turing Triage Test" evaluates whether and to what extent the continued existence of a computer can be comparable to another human being in what is arguable a highly constrained and somewhat artificial situation of life and death. In other words, it can be said that a computer has achieved moral standing that is at least on par with that of another human being, when it is possible that one could in fact choose the continued existence of the computer over that of another human individual. Although Sparrow's characterization appears to make the moral status of the machine dependent on its ability to simulate human characteristics, he is careful to avoid the trap of simple anthropomorphism. The issue, he contends, is whether "intelligent computers might achieve the status of moral persons"[35] and, following the example provided by animal ethicists like Peter Singer, Sparrow argues that the category "personhood" must be understood apart from the concept of the human. "Whatever it is that makes human beings morally significant must be something that could conceivably be possessed by other entities. To restrict personhood to human beings is to commit the error of chauvinism or 'speciesism.'"[36] Despite this important qualification, however, the definition of "personhood," which Sparrow admits is itself open to considerable equivocation, is something that is dependent upon and abstracted from human experience. Even if "moral personhood" is minimally characterized "as a capacity to experience pleasure and pain,"[37] what constitutes "pleasure" and "pain" is derived from and defined according to distinctly human experiences and values. Consequently, Sparrow's dependency on the innovations of animal rights philosophy, leads to his inheriting one of its fundamental problems, namely, extending what are essentially human values and interests to the animal-other does not necessarily contest but often reaffirms anthropocentrism. "One of the central ironies of animal rights philosophy," Wolfe writes, "is that its philosophical frame remains an essentially humanist one in its most important philosophers (utilitarianism in Peter Singer, neo-Kantianism in Tom Regan), thus effacing the very difference of the animal other that animal rights sought to respect in the first place."[38] Android Ethics, then, introduces an important and fundamental challenge to the anthropocentric privilege that has organized traditional forms of moral reasoning. At the same time, however, it does so by leveraging and redeploying essentially humanist values that unfortunately threaten to undermine its own innovations.

Like Sparrow's "Android Ethics," Luciano Floridi's "Information Ethics" is formulated in response to questions concerning the proper limits of moral consideration. According to Floridi's analysis, "any action, whether morally loaded or not, has the logical structure

---

[32] Robert Sparrow, "The Turing Triage Test," *Ethics and Information Technology* 6.4 (December 2004): 203.
[33] Ibid., 204.
[34] Ibid.

[35] Ibid.
[36] Ibid., 207.
[37] Ibid.
[38] Wolfe, xii.

of a binary relation between an agent and a patient."[39] Standard or classic forms of ethics have been exclusively concerned with either the character of the agent, as in virtue ethics, or the actions that are performed by the agent, as in consequentialism, contractualism, and deontologism. For this reason, Floridi concludes, classic ethical theories have been "inevitably anthropocentric" in focus, and "take only a relative interest in the patient," or what he also refers to as the "receiver" or "victim."[40] This philosophical status quo has been, Floridi suggests, recently challenged by animal and environmental ethics, both of which "attempt to develop a patient-oriented ethics in which the 'patient' may be not only a human being, but also any form of life."[41] However innovative this alteration has been, Floridi finds it to be insufficient for a truly universal and impartial ethics. "Even Bioethics and Environmental Ethics," he argues, "fail to achieve a level of complete universality and impartiality, because they are still biased against what is inanimate, lifeless, or merely possible (even Land Ethics is biased against technology and artefacts, for example). From their perspective, only what is alive deserves to be considered as a proper center of moral claims, no matter how minimal, so a whole universe escapes their attention."[42] For Floridi, therefore, bioethics and environmental ethics represents something of an incomplete innovation in moral philosophy. They have, on the one hand, successfully challenged the anthropocentric tradition by articulating a more universal form of ethics that not only shifts attention to the patient but also expands who or what qualifies for inclusion as a patient. At the same time, however, both remain ethically biased insofar as they substitute biocentrism for the customary anthropocentrism. Consequently, Floridi endeavors to take the innovations introduced by bioethics and environmental ethics one step further. He retains the patient-orient approach but "lowers the condition that needs to be satisfied, in order to qualify as a center of moral concern, to the minimal common factor shared by any entity"[43] whether animate, inanimate, or otherwise. For Floridi this common denominator is informational and, for this reason, he gives this thesis the name "Information Ethics" or IE.

> From an IE perspective, the ethical discourse now comes to concern information as such, that is not just all persons, their cultivation, well-being and social

interactions, not just animals, plants and their proper natural life, but also anything that exists, from paintings and books to stars and stones; anything that may or will exist, like future generations; and anything that was but is no more, like our ancestors. Unlike other non-standard ethics, IE is more impartial and universal – or one may say less ethically biased – because it brings to ultimate completion the process of enlargement of the concept of what may count as a center of information, no matter whether physically implemented or not.[44]

Following the innovations of bio- and environmental ethics, Floridi expands the scope of moral philosophy by altering its focus and lowering the threshold for inclusion. What makes someone or something a moral patient, deserving of some level of ethical consideration (no matter how minimal), is that it exists as a coherent body of information. This is a promising proposal, because it not only is able to incorporate a wider range of possible subjects (living organisms, organizations, works of art, machines, historical entities, etc.) but expands the scope of ethical thinking to include those others who have been, for one reason or another, traditionally excluded from moral consideration. In fact, in focusing attention on the patient of the action, Floridi's proposal comes as close as any analytic philosopher has to approximating the "ethics of otherness" that has been the hallmark of continental thinkers like Levinas. Despite this, however, IE still runs up against significant structural and philosophical difficulties. First, in shifting emphasis from an agent-oriented to a patient-oriented ethics, Floridi simply inverts the two terms of a traditional binary structure. If classic ethical thinking has been organized, for better or worse, by an interest in the character and actions of the agent at the expense of the patient, IE endeavors, following the innovations modeled by bioethics, to reorient things by placing emphasis on the depreciated term. This maneuver is, quite literally, a revolutionary proposal, because it inverts or "turns over" the traditional arrangement. Inversion, however, is rarely in and by itself a satisfactory mode of critical intervention. As Nietzsche, Levinas, and other post-structuralists have pointed out, the inversion of a binary opposition actually does little or nothing to challenge the fundamental structure of the system in question. In fact, inversion preserves and maintains the traditional structure, albeit in an inverted form. The effect of this on IE is registered by Kenneth Einar Himma, who, in an assessment of Floridi's argument, demonstrates that a concern for the patient is nothing more than the flip-side of good-old, agent-oriented ethics. "To say that an

---

[39] Luciano Floridi, "Information Ethics: On the Philosophical Foundation of Computer Ethics," *Ethics and Information Technology* 1.1 (March 1999): 41.
[40] Ibid., 41 and 42.
[41] Ibid., 42.
[42] Ibid., 43.
[43] Ibid.

[44] Ibid.

entity X has moral standing (i.e., is a moral patient) is, at bottom, simply to say that it is possible for a moral agent to commit a wrong against X. Thus, X has moral standing if and only if (1) some moral agent has at least one duty regarding the treatment of X and (2) that duty is owed to X."[45] According to Himma's analysis, IE's patient oriented ethics is not that different from traditional ethics, it simply looks at the agent/patient couple from the other side. Levinas, by contrast, does in fact introduce something entirely different. Instead of simply flipping the relative positions occupied by the agent and patient in the binary structure that has characterized traditional forms of moral theorizing, he considers the ethical experience of an Other that exceeds and remains exterior to these logical distinctions. "Experience, the idea of infinity," Levinas writes, "occurs in the relationship with the other. The idea of infinity is the social relationship. This relationship consists in approaching an absolutely exterior being. The infinity of this being, which one can therefore not contain, guarantees and constitutes this exteriority. It is not equivalent to the distance between a subject and an object."[46] In this way, Levinas's "ethics of otherness" is concerned with an Other who is not defined, as Floridi's patient is, as the mere flip-side of the agent or self-same; it is entirely and radically otherwise.

Second, IE not only alters the orientation of ethics but also enlarges its scope by reducing the minimum requirements for inclusion. It replaces both the exclusive anthropocentric and biocentric theories with an "ontocentric" one, which is, by comparison, much more inclusive. In doing so, however, IE simply replaces one form of centrism with another. This is, as Levinas points out, really nothing different; it is more of the same. "Western philosophy has most often been an ontology: a reduction of the other to the same by interposition of a middle or neutral term that ensures the comprehension of being."[47] According to Levinas's analysis, the standard operating procedure of western philosophy has been the reduction of difference. In fact, philosophy has, at least since the time of Aristotle, usually explained and dealt with difference by finding below and behind apparent variety some common denominator that is and remains irreducibly the same. Anthropocentric ethics, for example, posits a common humanity that underlies and substantiates the perceived differences in race, gender, ethnicity, class, etc. Likewise, biocentric ethics assumes that there is a common value in life itself, which subtends all forms of available biological diversity. And in the ontocentric theory of IE, it is being, the very matter of ontology itself, that underlies and supports all apparent differentiation. As Himma describes it, "every existing entity, whether sentient or non-sentient, living or non-living, natural or artificial, has some minimal moral worth…in virtue of its existence."[48] But as Levinas argues, this desire to articulate a universal, common element effectively reduces the difference of the other to what is ostensibly the same. "Perceived in this way," Levinas writes, "philosophy would be engaged in reducing to the same all that is opposed to it as other."[49] In taking an ontocentric approach, therefore, IE reduces all difference to a minimal common factor that is supposedly shared by any and all entities. Although this approach provides for a more inclusive kind of "centrism," it still utilizes a centrist approach and, as such, necessarily reduces difference to some preselected common denominator. None of this, however, should be taken to mean that Levinas simply trumps Floridi, which would ignore the fact that Floridi's work questions and complicates Levinas's adherence to humanism. What it does mean is that the innovation that has been ascribed to IE may not be as unconventional and different from the mainstream of moral philosophy as was initially advertised.

Otherwise

> In the history of the United States, gradually more and more beings have been granted the same rights that others possessed and we've become a more ethical society as a result. Ethicists are currently struggling with the question of whether at least some higher animals should have rights, and the status of human fetuses has been debated as well. On the horizon looms the question of whether intelligent machines should have moral standing.[50]
> – Susan Leigh Anderson

As is evident from even a cursory review of the history of philosophy, ethics has been an exclusive undertaking. For most of us, it is not news that moral philosophy has been and, in many cases, continues to be organized around the human subject. The problem in this is, of course, the fact that the concept of the *human*

---

[45] Kenneth Einar Himma, "There's Something About Mary: The Moral Value of Things *qua* Information Objects," *Ethics and Information Technology* 6.3 (September 2004): 145.

[46] Emmanuel Levinas, *Collected Philosophical Papers*, trans. Alphonso Lingis (Dordrecht: Martinus Nijhoff Publishers, 1987), 54–55.

[47] Levinas, *Totality*, 43.

[48] Himma, 145.

[49] Levinas, *Collected*, 43.

[50] Susan Leigh Anderson, "Asimov's 'Three Laws of Robotics' and Machine Metaethics," *AI and Society* (forthcoming): 3. Preprint available at http://www.springerlink.com/content/771k1181268772p1

has been arbitrary, flexible, and not altogether consistent. At different times, the membership criteria for club-anthropos has been defined in such a way as to not only exclude but justify the exclusion of others, e.g., barbarians, women, Jews, people of color, etc. As membership in the club has slowly and not without considerable resistance been extended to these excluded populations, there have remained other, apparently more fundamental, exclusions, most notably that of the animal and the machine. And even the recent innovations introduced under the banner of animal rights, although securing some form of access by non-human animals, has continued to exclude the machine. This exclusion is theoretically unjustified. Because the animal and machine, at least since the time of Descartes, share a common form of alterity, the one cannot be admitted without also opening the door to the other. So despite all the innovations in moral philosophy by which both human and non-human others have been extended some claim to moral standing, the exclusion of the machine appears to be the last socially accepted moral prejudice.

The exclusion of the machine from the proper domain of ethics is certainly an ethical problem. But inclusion, as its mere flip-side and dialectical other, appears to be no less problematic. Despite the recent political and intellectual cache that has accrued to the word, "inclusion" is not without significant ethical complications and consequences. The inclusion of the other, whether another human being, the animal, the environment, the machine, or something else, always and inevitably runs up against the same difficulty, namely the reduction of difference to the same. In order to extend the boundaries of moral philosophy to traditionally excluded others, philosophers have argued for progressively more inclusive definitions of what qualifies someone or something for ethical consideration. That is, they have continually shifted the level of abstraction by which two different things come to be recognized as essentially the same and therefore deserving of each other's respect. Anthropocentrism, for example, situates the human at the center of ethics and admits into moral consideration anyone who is able to meet the basic criteria of what has been decided to comprise the human. Animiocentrism focuses attention on the animal and extends consideration to any organism that meets the defining criteria of animality. Biocentrism goes one step further in the process of abstraction; it defines life as the common denominator and admits into consideration anything and everything that can be said to be alive. And ontocentrism completes the progression by incorporating into moral consideration anything that actually exits, had existed, or potentially exists. All of these innovations, despite their differences in focus, employ a similar maneuver. That is,

they redefine the center of moral consideration in order to describe progressively wider circles that come to encompass a greater number of possible participants. Although there are and will continue to be considerable debates about who or what should define the center and who or what is or is not included, this debate is not the problem. The problem rests in the strategy itself. In taking a centrist approach, these different ethical theories endeavor to identify what is essentially the same in a phenomenal diversity of individuals. Consequently, they include others by effectively stripping away and reducing their differences. This approach although having the appearance of being increasingly more inclusive, immediately effaces the unique alterity of others and turns them into more of the same, instituting what Slavoj Žižek calls the structure of the mobius band: "At the very heart of Otherness, we encounter the other side of the Same."[51] This is, as Levinas, Cary Wolfe, and others have pointed out, a considerable moral problem in that it effectively eradicates the very differences one had sought to take into account and respect. In making this argument, however, it should be noted that the criticism has itself employed what it criticizes. In focusing attention on what is essentially the same in these various forms of moral centrism, the analysis does exactly what it charges – it identifies a common feature that underlies apparent diversity and effectively reduces a multiplicity of differences to what is the same. Pointing this out, however, does not invalidate the conclusion but demonstrates, not only in what is said but also in what is done, the questionable operations that are already involved in any attempt at articulating inclusion.

Exclusion is a problem because it calls attention to and fixates on what is different despite potential similarities. Inclusion is a problem, because it emphasizes similarities at the expense of respecting important differences. Consequently, the one is the inverse of the other, or, to put it in colloquial terms, they are two sides of one coin. As long as ethical debate and innovation remains involved with and structured by these two possibilities little or nothing will change. Exclusion will be identified and challenged, as it has been in the discourses of animal rights, bioethics, and information ethics by calls for greater inclusivity and ethical theories that are able to accommodate others. At the same time, inclusion will be challenged, as it has in critical responses to the project of animal rights, for its reduction of difference and reappropriation of the otherness that it had sought to respect and accommodate in the first place. What is need, therefore, is a third alternative that does not simply oppose exclusion to inclusion or vice versa. Fortunately examples of this

---

[51] Slavoj Žižek, *The Plague of the Fantasies* (New York: Verso, 1997), 161.

kind of alternative approach can already be found in both philosophy and the field of AI.

In *The Inclusion of the Other*, for example, Jürgen Habermas (1998) proposes what amounts to a non-reductive universalism that is designed to be highly sensitive to differences: "The equal respect for everyone else demanded by a moral universalism sensitive to difference thus takes the form of a *nonleveling* and *nonappropriating inclusion* of the other in his *otherness*."[52] In proposing "a nonleveling and nonappropriating inclusion," Habermas attempts to identify an alternative to the terms that have traditionally structured moral philosophy. This "third term," which can only be articulated in language by using what appears to be contradictory predicates, exceeds and intentionally violates the either/or logic of inclusion/exclusion. A similar maneuver has been proposed in the writings of Levinas. In fact, it is Levinas who, more so than any other thinker in the western tradition, provides what is arguably the most elaborate and sustained consideration of this problem. Levinasian philosophy is not only critical of the traditional tropes and traps of western ontology but proposes an ethics of radical otherness that deliberately resists and interrupts the metaphysical gesture par excellence, that is, the reduction of difference to the same. Despite these promising innovations, however, one needs to be aware of and to work against the persistent and irreducible humanism that has been shown to pervade and underlie the work of Levinas, and those others who, following his example, endeavor to address themselves to otherness. We must, therefore, as Jacques Derrida once wrote of Georges Bataille's exceedingly careful engagement with the thought of Hegel, follow Levinas to the end, "to the point of agreeing with him against himself" and of wresting his discoveries from the limited interpretations that he had provided.[53] One way to facilitate this process is to encourage communication between and to work deliberately in excess of the main division that currently organizes philosophical discourse – the decision that, for better or worse, parses the discipline into analytic and continental varieties. On the one hand, analytic theorist working in both animal rights philosophy and information ethics have provided compelling arguments by which to challenge and to undermine the tradition of anthropocentrism that has been dominant in ethics. They have, however, done so at the expense of erecting other, no less problematic centrisms and, in the process, have often repeated in practice what they had

opposed in theory. Continental theorists, on the other hand, have effectively criticized the philosophy of the same that is at the center of all centrisms and have successfully described the structure and exigencies of an ethics that is oriented otherwise. At the same time, however, they have done so on the basis and in the name of what remains an essentially unquestioned validation of the human and the traditions of humanism. If ever there was a time and a reason for the one side to take seriously the innovations of the other, it is in the face of the machine and in response to what can now be called "the machine question."

Attention to another form of otherness that is not simply more of the same has also been introduced and exemplified in recent attempts to account for machine intelligence. According to Rodney Brooks, "the most important change in AI happened in the 1980s when some people realized that the model of reasoning used in AI was very different from what happens inside the heads of people, very different at any level of abstraction used for the descriptions. Such differences do not invalidate the nonhuman approaches – airplanes are good examples of very useful machines that operate very differently from the way real birds operate. But the realization in AI opened up new ways of doing things and new avenues to go down…"[54] As long as the project of AI is pursued, as it had been for many years since its inception in the 1950's, as an attempt to get computers to think just like human beings, then it is an undertaking that will continually fall short of expectations. There is, in fact, no machine that can "think" the same way the human entity thinks and all attempts to get machines to simulate the activity of human thought processes, no matter what level of abstraction is utilized, have lead to considerable frustration or outright failure. If, however, one recognizes, as many AI researchers have since the 1980's, that machine intelligence may take place and be organized completely otherwise, then a successful "thinking machine" is not just possible but may already be extant. Following this precedent, we can say that just as there are other orders of intelligence that need to be accounted for in ways that do not simply identify similarities to ourselves, there may also be alternative moral subjects and ethical capabilities that need to be understood and addressed otherwise.

In the end, therefore, it can be said that the machine constitutes a significant challenge to the customary standards and practices of moral reasoning. This challenge, which I have identified, following the precedent of an earlier innovation, with the moniker *the machine question*, is fundamental and irreducible. And

---

[52] Jürgen Habermas, *The Inclusion of the Other: Studies in Political Theory*, trans. Ciaran Cronin et al. (Cambridge, MA: MIT Press, 1998), 40.

[53] Jacques Derrida, *Writing and Difference*, trans. Alan Bass (Chicago: University of Chicago Press, 1978), 260.

---

[54] Rodney Brooks. Comment in "AI's Greatest Trends and Controversies," by Marti A. Hearst and Haym Hirsh. *IEEE – Intelligent Systems* 15.1 (January/February 2002): 8–17.

this is what makes it both theoretically interesting and ethically important. The machine question is not, for instance, simply interested in identifying and documenting the exclusion of the machine from the domain of ethics. Nor is it concerned, like the recent innovations introduced in animal, machine, android, or information ethics, with formulating criteria and strategies for its inclusion. The machine question is certainly concerned with both these aspects, but it also and necessarily involves more. In particular, it is and must be concerned with the ethical complications and side-effects that are imposed by the deployment and use of the binary pairing of inclusion/exclusion in these and all other moral considerations. The machine question, therefore, puts in question the entire edifice of ethics and the mechanisms by which moral philosophy has, until now, successfully differentiated and distinguished what comes to be included and what has been excluded. To paraphrase Floridi, and to agree with his analysis in excess of the restricted interpretations he gives it, the machine question not only adds interesting new dimensions to old problems, but leads us to rethink, methodologically, the very grounds on which our ethical positions are based.[55] The machine question,[56] then, not only necessitates responsiveness to others and other forms of otherness but requires modes of oper-

ation that are oriented and organized otherwise. In other words, it is in the confrontation with the alterity of the machine and the fundamental challenges it introduces that moral philosophy comes to experience its own structural limits and the need for thinking otherwise.

---

[55] Luciano Floridi, "Information Ethics, its Nature and Scope," in *Moral Philosophy and Information Technology*, eds. Jeroen van den Hoven and John Weckert (Cambridge: Cambridge University Press, 2006). Preprint available at http://www.wolfson.ox.ac.uk/~floridi/papers.htm, p. 7.

[56] Clearly we will, at some point, want to advance and to entertain answers to the machine question. Such answers will not only need to decide, among other things, what Susan Leigh Anderson and others have referred to as the "moral standing of intelligent machines" (2007: 2) but will also need to evaluate, as is evident from the above critique of machine, android, and information ethics, the moral presumptions and consequences necessarily instituted by all such decisions. Because this is both complex and, as Anderson adds, not easy to accomplish (2007: 2), I suggest, at least for the time being, a much more modest and restrained undertaking. That is, I propose that we take some time to articulate the question and to attend to its metaphysical contours, moral assumptions, and epistemological structures. In doing so, I follow the precedent of Martin Heidegger, who points out, at the beginning of *Being and Time* (New York: Harper & Row, 1962) that questioning is never a neutral or uninformed activity. The questions that one asks already decide, well in advance of anything that begins to look like an answer, what can asked, who can be interrogated, and what kinds of responses will count as appropriate. And it is for this reason that I offer the preceding as a small but hopefully significant contribution to our admittedly nascent understanding of the machine question.

References

G. Agamben, *The Open: Man and Animal*. Stanford University Press, Stanford, CA, 2004.

S.L. Anderson. Asimov's 'Three Laws of Robotics' and Machine Metaethics. *AI and Society*, 3, 2007. Preprint available at http://www.springerlink.com/content/771k1181268772p1.

M. Anderson and S.L. Anderson. The Status of Machine Ethics: A Report from the AAAI Symposium. *Mind and Machines*, 17: 1–10, 2007. Preprint available at http://www.springerlink.com/content/e502572456857542.

M. Anderson, S.L. Anderson, and C. Armen. Toward Machine Ethics. American Association for Artificial Intelligence – The Nineteenth National Conference on Artificial Intelligence, 25–29 July 2004, San Jose, CA.

I. Asimov, *Robot*. Bantam Books, New York, 1991.

J. Derrida. *Writing and Difference* (trans. Alan Bass), p. 260. University of Chicago Press, Chicago, 1978.

J. Derrida. And Say the Animal Responded. In C. Wolfe, editor, *Zoontologies: The Question of the Animal* (trans. David Willis), p. 121. University of Minnesota Press, Minneapolis, MN, 2003.

R. Descartes. Discourse on Method. In *Descartes: Selected Philosophical Writings* (trans. and editors J. Cottingham, R. Stoothoff and D. Murdoch), p. 44. Cambridge University Press, Cambridge, 1988.

R.F. Fidler, *Mediamorphosis: Understanding New Media*. Pine Forge Press, Thousand Oaks, CA, 1997.

L. Floridi. Information Ethics: On the Philosophical Foundation of Computer Ethics. *Ethics and Information Technology*, 1(1): 41, 1999.

L. Floridi. Information Ethics, its Nature and Scope. In J. van den Hoven and J. Weckert, editors, *Moral Philosophy and Information Technology*, p. 7. Cambridge University Press, Cambridge, 2006. Preprint available at http://www.wolfson.ox.ac.uk/~floridi/papers.htm.

J. Habermas. *The Inclusion of the Other: Studies in Political Theory* (trans. Ciaran Cronin et al.), p. 40. MIT Press, Cambridge, MA, 1998.

J.S. Hall. Ethics for Machines. *KurzweilAI.net* (5 July 2001), http://www.kurzweilai.net/articles/art0218.html.

D. Haraway, *Simian, Cyborgs and Women: The Reinvention of Nature*. Routledge, New York, 1991 151–152.

M. Heidegger. *Die Grundbegriffe der Metaphysik: Welt – Endlichkeit – Einsamkeit*. V. Klostermann, Frankfurt am Main, 1983.

K.E. Himma. There's Something About Mary: The Moral Value of Things *qua* Information Objects. *Ethics and Information Technology*, 6(3): 145, 2004.

R. Holeton, editors, *Composing Cyberspace: Identity, Community, and Knowledge in the Electronic Age*. McGraw Hill, New York, 1998.

B. Johnson. Translator's Introduction. In J. Derrida, editors, *Disseminations*, pp. xv–. University of Chicago Press, Chicago, 1981.

I. Kant. *Critique of Practical Reason* (trans. Lewis W. Beck), p. 17. Macmillan, New York, 1985.

S. Kiesler, editors, *Culture of the Internet*. Lawrence Erlbaum Associates, Mahwah, NJ, 1997.

D.F. Krell. *Daimon Life: Heidegger and Life Philosophy*. Indiana University Press, Bloomington, IN, 1992.

E. Levinas. *Totality and Infinity* (trans. Alphonso Lingis). Duquesne University Press, Pittsburgh, PA, 1969.

E. Levinas. *Otherwise than Being or Beyond Essence* (trans. Alphonso Lingis). Martinus Nijhoff Publishers, The Hague, 1981.

E. Levinas. *Collected Philosophical Papers* (trans. Alphonso Lingis), pp. 54–55. Martinus Nijhoff Publishers, Dordrecht, 1987.

J.-F. Lyotard. *The Inhuman: Reflections on Time* (trans. Geoffrey Bennington and Rachel Bowlby), p. 1. Stanford University Press, Stanford, CA, 1991.

W.J. Mitchell, *City of Bits: Space, Place, and the Infobahn*. MIT Press, Cambridge, 1995.

J.T. Nealon, *Alterity Politics: Ethics and Performative Subjectivity*. Duke University Press, Durham, NC, 1998 71.

E. Ratliff. The Crusade Against Evolution. *Wired* 12.10: 156–161, 2004.

T. Regan. In P. Steeves, editor, *Animal Others: On Ethics, Ontology, and Animal Life*, p. xii. State University of New York Press, Albany, NY, 1999.

D. Saco, *Cybering Democracy: Public Space and the Internet*. University of Minnesota Press, Minneapolis, MN, 2002.

R. Sparrow. The Turing Triage Test. *Ethics and Information Technology*, 6(4): 203, 2004.

P. Steiner. Dog cartoon. *The New Yorker*, 61, 5 July 1993.

A.R. Stone, *The War of Desire and Technology at the Close of the Mechanical Age*. MIT Press, Cambridge, MA, 1995.

S. Turkle, *Life on the Screen: Identity in the Age of the Internet*. Simon & Schuster, New York, 1995.

K. Warwick. Cyborg Morals, Cyborg Values, Cyborg Ethics. *Ethics and Information Technology*, 5(3), 2003.

C. Wolfe, editor. *Zoontologies: The Question of the Animal*, pp. x–xi. University of Minnesota Press, Minneapolis, MN, 2003.

S. Žižek, *The Plague of the Fantasies*. Verso, New York, 1997 161.