This is the author's accepted manuscript of a book chapter that has been published in Zizi Papacharissi (Ed.), *A Networked Self: Human Augmentics, Artificial Intelligence, Sentience* (pp. 51-68). Routledge, 2018. Please do not circulate.

Other Things:
AI, Robots and Society

David J. Gunkel – Northern Illinois University (USA)

We are it seems in the midst of a robot apocalypse. The invasion, however, does not look like what we have been programmed to expect from decades of science fiction literature and film. It occurs not as a spectacular catastrophe involving a marauding army of alien machines descending from the heavens with weapons of immeasurable power. Instead, it takes place, and is already taking place, in ways that look more like the fall of Rome than *Battlestar Galactica*, with machines of various configurations and capabilities slowly but surely coming to take up increasingly important and influential positions in everyday social reality. "The idea that we humans would one day share the Earth with a rival intelligence," Philip Hingston (2014) writes, "is as old as science fiction. That day is speeding toward us. Our rivals (or will they be our companions?) will not come from another galaxy, but out of our own strivings and imaginings. The bots are coming: chatbots, robots, gamebots."

And the robots are not just coming. They are already here. In fact, our communication and information networks are overrun, if not already run, by machines. It is now estimated that over 50% of online traffic is machine generated and consumed (Zeifman 2017). This will only increase with the Internet of things (IoT), which is expected to support over 26 billion interactive and connected devices by 2020 (by way of comparison, the current human population of planet earth is estimated to be 7.4 billion) (Gartner 2013). We have therefore already achieved and live in that future Norbert Wiener (1950) had predicted at the beginning of *The Human Use of Human Beings: Cybernetics and Society*: "It is the thesis of this book that society can only be understood through a study of the messages and the communication facilities which belong to it; and that in the future development of these messages and communication facilities, messages

between man and machines, between machines and man, and between machine and machine, are destined to play an ever-increasing part" (p. 16).

What matters most in the face of this machine incursion is not resistance—insofar as resistance is already futile—but how we decide to make sense of and respond to the new social opportunities or challenges that these things make available to us. The investigation of this matter will proceed through three steps or movements. The first part will critically reevaluate the way we typically situate and make sense of things. It will therefore target and reconsider the instrumental theory, which characterizes things, and technological artifacts in particular, as nothing more than tools serving human interests and objectives. The second will investigate the opportunities and challenges that recent developments with artificial intelligence, learning algorithms, and social robots pose to this standard default understanding. These other kinds of things challenge and exceed the conceptual boundaries of the instrumental theory and ask us to reassess who or what is (or can be) a legitimate social subject. Finally, and by way of conclusion, the third part will draw out the consequences of this material, explicating what this development means for us, the other entities with which we communicate and interact, and the new social situations and circumstances that are beginning to define life in the 21st century.

1 Standard Operating Presumptions

There is, it seems, nothing particularly interesting or extraordinary about things. We all know what things are; we deal with them every day. But as Martin Heidegger (1962) pointed out, this immediacy and proximity is precisely the problem. Marshall McLuhan and Quentin Fiore (2001) cleverly explained it this way: "one thing about which fish know exactly nothing is water" (p. 175). Like fish that cannot perceive the water in which they live and operate, we are, Heidegger argues, often unable to see the things that are closest to us and comprise the very milieu of our everyday existence. In response to this, Heidegger commits considerable effort to investigating what things are and why things seem to be more difficult than they initially appear. In fact, "the question of things," is one of the principal concerns and an organizing principles of Heidegger's ontological project (Benso, 2000, p. 59), and this concern with things begins right at the beginning of his 1927 magnum opus, *Being and Time*: "The Greeks had an appropriate term for 'Things': πράγματα [*pragmata*]—that is to say, that which one has to do with in one's concernful dealings (πραξις). But ontologically, the specific 'pragmatic' character of the

πράγματα is just what the Greeks left in obscurity; they thought of these 'proximally' as 'mere Things'. We shall call those entities which we encounter in concern '*equipment*' [*Zeug*]" (Heidegger, 1962, p. 96-97).

According to Heidegger's analysis, things are not, at least not initially, experienced as mere entities out there in the world. They are always pragmatically situated and characterized in terms of our involvements and interactions with the world in which we live. For this reason, things are first and foremost revealed as "equipment," which are useful for our endeavors and objectives. "The ontological status or the kind of being that belongs to such equipment," Heidegger (1962) explains, "is primarily exhibited as 'ready-to-hand' or *Zuhandenheit*, meaning that some-thing becomes what it is or acquires its properly 'thingly character' when we use it for some particular purpose" (p. 98). According to Heidegger, then, the fundamental ontological status, or mode of being, that belongs to things is primarily exhibited as "ready-to-hand," meaning that something becomes what it is or acquires its properly "thingly character" in coming to be put to use for some particular purpose. A hammer, one of Heidegger's principal examples, is for building a house to shelter us from the elements; a pen is for writing an essay like this; a shoe is designed to support the activity of walking. Everything is what it is in having a "for which" or a destination to which it is always and already referred. Everything therefore is primarily revealed as being a tool or an instrument that is useful for our purposes, needs, and projects.[1]

This mode of existence—what Graham Harman (2002) calls "tool-being"—applies not just to human artifacts, like hammers, pens, and shoes. It also describes the basic ontological condition of natural objects, which are, as Heidegger (1962) explains, discovered in the process of being put to use: "The wood is a forest of timber, the mountain a quarry of rock, the river is water-power, the wind is wind 'in the sails'" (p. 100). Everything therefore exists and become what it is insofar as it is useful for some humanly defined purpose. Things are not just out there in a kind of raw and naked state but come to be what they are in terms of how they are already put to work and used as equipment for living. And this is what makes things difficult to see or perceive. Whatever is ready-to-hand is essentially transparent, unremarkable, and even invisible. "The peculiarity," Heidegger (1962) writes, "of what is proximally ready-to-hand is that, in its readiness-to-hand, it must as it were, withdraw in order to be ready-to-hand quite authentically. That with which our everyday dealings proximally dwell is not the tools themselves. On the

contrary, that with which we concern ourselves primarily is the work" (p. 99). Or as Michael Zimmerman (1990) explains by way of Heidegger's hammer, "In hammering away at the sole of a shoe, the cobbler *does not notice the hammer*. Instead, the tool is in effect transparent as an extension of his hand…For tools to work right, they must be 'invisible,' in the sense that they disappear in favor of the work being done" (p. 139).

This understanding of things can be correlated with the "instrumental theory of technology," which Heidegger subsequently addresses in *The Question Concerning Technology* (1970). As Andrew Feenberg (1991) has summarized it, "the instrumentalist theory offers the most widely accepted view of technology. It is based on the common sense idea that technologies are 'tools' standing ready to serve the purposes of users" (p. 5). And because a tool or an instrument "is deemed 'neutral,' without valuative content of its own" a technological thing is evaluated not in and of itself, but on the basis of the particular employments that have been operationalized by its human designer, manufacturer, or user. Following from this, technical devices, no matter how sophisticated or autonomous they appear or are designed to be, are typically not considered the responsible agent of actions that are performed with or through them. "Morality, "as J. Storrs Hall (2001) points out, "rests on human shoulders, and if machines changed the ease with which things were done, they did not change responsibility for doing them. People have always been the only 'moral agents'" (p. 2). To put it in colloquial terms (which nevertheless draw on and point back to Heidegger's example of the hammer): "It is a poor carpenter who blames his tools."

This way of thinking not only sounds level-headed and reasonable, it is one of the standard assumptions deployed in the field of technology and computer ethics. According to Deborah Johnson's (1985) formulation, "computer ethics turns out to be the study of human beings and society—our goals and values, our norms of behavior, the way we organize ourselves and assign rights and responsibilities, and so on" (p. 6). Computers, she recognizes, often "instrumentalize" these human values and behaviors in innovative and challenging ways, but the bottom-line is and remains the way human agents design and use (or misuse) such technology. Understood in this way, computer systems, no matter how automatic, independent, or seemingly intelligent they may become, "are not and can never be (autonomous, independent) moral agents" (Johnson, 2006, p. 203). They will, like all other things, always be instruments of human value, decision making, and action.

2 Other Kinds of Things

This instrumentalist way of thinking not only sounds reasonable, it is obviously useful. It is, one might say, instrumental for parsing and responding to questions concerning proper conduct and social responsibility in the age of increasingly complex technological devices and systems. And it has a distinct advantage in that it locates accountability in a widely-accepted and seemingly intuitive subject position, in human decision making and action. At the same time, however, this particular formulation also has significant theoretical and practical limitations, especially as it applies (or not) to recent innovations. Let's consider three examples that not only complicate the operative assumptions and consequences of the instrumental theory but require new ways of perceiving and theorizing the social challenges and opportunities of things.

2.1 Things that Talk

From the beginning, it is communication—and specifically, a tightly constrained form of conversational interpersonal dialogue—that provides the field of artificial intelligence (AI) with its definitive characterization and test case. This is immediately evident in the agenda-setting paper that is credited with defining machine intelligence, Alan Turing's "Computing Machinery and Intelligence," which was first published in the journal *Mind* in 1950. Although the term "artificial intelligence" is a product of the Dartmouth Conference of 1956, it is Turing's seminal paper and the "game of imitation" that it describes—what is now routinely called "the Turing Test"—that defines and characterizes the field. "The idea of the test," Turing (2004) explained in a BBC interview from 1952, "is that the machine has to try and pretend to be a man, by answering questions put to it, and it will only pass if the pretense is reasonably convincing. A considerable proportion of a jury, who should not be experts about machines, must be taken in by the pretense. They aren't allowed to see the machine itself—that would make it too easy. So the machine is kept in a faraway room and the jury are allowed to ask it questions, which are transmitted through to it" (p. 495). According to Turing's stipulations, if a machine is capable of successfully simulating a human being in communicative interactions to such an extent that human interlocutors (or "a jury" as Turing calls them in the 1952 interview) cannot tell whether they are talking with a machine or another human being, then that device would need to be considered intelligent (Gunkel 2012b).

At the time that Turing published the paper proposing this test-case, he estimated that the tipping point—the point at which a machine would be able to successfully play the game of imitation—was at least half-a-century in the future. "I believe that in about fifty years' time it will be possible to programme computers, with a storage capacity of about 10^9, to make them play the imitation game so well that an average interrogator will not have more than 70 per cent chance of making the right identification after five minutes of questioning" (Turing, 1999, p. 44). It did not take that long. Already in 1966 Joseph Weizenbaum demonstrated a simple natural language processing (NLP) application that was able to converse with human interrogators in such a way as to appear to be another person. ELIZA, as the application was called, was what we now recognize as a "chatterbot." This proto-chatterbot[2] was actually a rather simple piece of programming, "consisting mainly of general methods for analyzing sentences and sentence fragments, locating so-called key words in texts, assembling sentence from fragments, and so on. It had, in other words, no built-in contextual framework of universe of discourse. This was supplied to it by a 'script.' In a sense ELIZA was an actress who commanded a set of techniques but who had nothing of her own to say" (Weizenbaum, 1976, p. 188). Despite this rather simple architecture, Weizenbaum's program demonstrated what Turing had initially predicted:

> ELIZA created the most remarkable illusion of having understood in the minds of many people who conversed with it. People who know very well that they were conversing with a machine soon forgot that fact, just as theatergoers, in the grip of suspended disbelief, soon forget that the action they are witnessing is not "real." This illusion was especially strong and most tenaciously clung to among people who know little or nothing about computers. They would often demand to be permitted to converse with the system in private, and would, after conversing with it for a time, insist, in spite of my explanations, that the machine really understood them (Weizenbaum, 1976, p. 189).

Since the debut of ELIZA, there have been numerous advancements in chatterbot design, and these devices now populate many of the online social spaces in which we live, work, and play. As a result of this proliferation, it is not uncommon for users to assume they are talking to another (human) person, when in fact they are just chatting up a chatterbot. This was the case for

Robert Epstein, a Harvard University PhD and former editor of *Psychology Today*, who fell in love with and had a four month online "affair" with a chatterbot (Epstein, 2007). This was possible not because the bot, that went by the name "Ivana," was somehow intelligent, but because the bot's conversational behavior was, in the words of Byron Reeves and Clifford Nass (1996), "close enough to human to encourage social responses" (p. 22). And this approximation is not necessarily "a feature of the sophistication of bot design, but of the low bandwidth communication of the online social space," where it is much easier to convincingly simulate a human agent (Mowbray, 2002, p. 2).

Despite this knowledge—despite educated, well-informed experts like Epstein (2007) who has openly admitted that "I know about such things and I should have certainly known better" (p. 17)—these software implementations can have adverse effects on both the user and the online communities in which they operate. To make matters worse (or perhaps more interesting) the problem is not something that is unique to amorous interpersonal relationships. "The rise of social bots," as Andrea Peterson (2013) accurately points out, "isn't just bad for love lives—it could have broader implications for our ability to trust the authenticity of nearly every interaction we have online" (p. 1). Case in point—national politics and democratic governance. In a study conducted during the 2016 US Presidential campaign, Alessandro Bessi and Emilio Ferrara (2016) found that "the presence of social media bots can indeed negatively affect democratic political discussion rather than improving it, which in turn can potentially alter public opinion and endanger the integrity of the Presidential election" (p. 1).

But who or what is culpable in these circumstances? The instrumental theory typically leads such questions back to the designer of the application, and this is precisely how Epstein (2007) made sense of his own experiences, blaming (or crediting) "a very smug, very anonymous computer programmer" who he assumes was located somewhere in Russia (p. 17). But things are already more complicated. Epstein is, at least, partially responsible for "using" the bot and deciding to converse with it, and the online community in which Epstein met Ivana is arguably responsible for permitting (perhaps even encouraging) such "deceptions" in the first place. For this reason, the assignment of culpability is not as simple as it might first appear to be. As Mowbray (2002) argues, interactions like this "show that a bot may cause harm to other users or to the community as a whole by the will of its programmers or other users, but that it also may cause harm through nobody's fault because of the combination of circumstances involving some

combination of its programming, the actions and mental or emotional states of human users who interact with it, behavior of other bots and of the environment, and the social economy of the community" (p. 4). Unlike artificial general intelligence (AGI), which would presumably occupy a subject position reasonably close to that of another human agent, these ostensibly mindless but very social things simply muddy the water (which is probably worse) by complicating and leaving undecided questions regarding agency and instrumentality.

2.2 Things that Think for Themselves

Standard chatterbot architecture, like many computer applications, depends on programmers coding explicit step-by-step instructions—ostensibly a set of nested conditional statements that are designed to respond to various kinds of input and machine states. In order to have ELIZA, or any other chatterbot, "talk" to a human user, human programmers need to anticipate everything that might be said to the bot and then code instructions to generate an appropriate response. If, for example, the user types "Hi, how are you." The application can be designed to identify this pattern of words and to respond with a pre-designated result, what Weizenbaum called a "script." Machine learning, however, provides an alternative approach to application design and development. "With machine learning," as *Wired* magazine explains, "programmers do not encode computers with instructions. They *train* them" (Tanz, 2016, p. 77). Although this alternative is nothing new—it was originally proposed and demonstrated by Arthur Samuel as early as 1956—it has recently gained popularity by way of some highly publicized events involving Google DeepMind's AlphaGo, which beat one of the most celebrated players of the notoriously difficult board game Go, and Microsoft's Twitterbot Tay.ai, which learned to become a hate spewing neo-Nazi racist after interacting with users on the Internet.

Both AlphaGo and Tay are AI systems using connectionist architecture. AlphaGo, as Google DeepMind (2015) explains "combines Monte-Carlo tree search with deep neural networks that have been trained by supervised learning, from human expert games, and by reinforcement learning from games of self-play." In other words, AlphaGo does not play the game of Go by following a set of cleverly designed moves described and defined in code by human programmers. The application is designed to formulate its own instructions from discovering patterns in existing data that has been assembled from games of expert human

players ("supervised learning") and from the trial-and-error experience of playing the game against itself ("reinforcement learning"). Although less is known about the exact inner workings of Tay, Microsoft explains that the system "has been built by mining relevant public data," i.e. training its neural networks on anonymized data obtained from social media, and was designed to evolve its behavior from interacting with users on social networks like Twitter, Kik, and GroupMe (Microsoft 2016a). What both systems have in common is that the engineers who designed and built them have no idea what these things will eventually do once they are in operation. As Thore Graepel, one of the creators of AlphaGo, has explained: "Although we have programmed this machine to play, we have no idea what moves it will come up with. Its moves are an emergent phenomenon from the training. We just create the data sets and the training algorithms. But the moves it then comes up with are out of our hands" (Metz, 2016, p. 1). Consequently, machine learning systems, like AlphaGo, are intentionally designed to do things that their programmers cannot anticipate or completely control. In other words, we now have autonomous (or at least semi-autonomous) things that in one way or another have "a mind of their own." And this is where things get interesting, especially when it comes to questions of social responsibility and behavior.

AlphaGo was designed to play Go, and it proved its ability by beating an expert human player. So who won? Who gets the accolade? Who actually beat the Go champion Lee Sedol? Following the dictates of the instrumental theory, actions undertaken with the computer would be attributed to the human programmers who initially designed the system and are capable of answering for what it does or does not do. But this explanation does not necessarily hold for an application like AlphaGo, which was deliberately created to do things that exceed the knowledge and control of its human designers. In fact, in most of the reporting on this landmark event, it is not Google or the engineers at DeepMind who are credited with the victory. It is AlphaGo. In published rankings, for instance, it is AlphaGo that is named as the number two player in the world (Go Ratings, 2016). Things get even more complicated with Tay, Microsoft's foul-mouthed teenage AI, when one asks the question: Who is responsible for Tay's bigoted comments on Twitter? According to the standard instrumentalist way of thinking, we would need to blame the programmers at Microsoft, who designed the application to be able to do these things. But the programmers obviously did not set out to create a racist Twitterbot. Tay

developed this reprehensible behavior by learning from interactions with human users on the Internet. So how did Microsoft answer for this? How did they explain things?

Initially a company spokesperson—in damage-control mode—sent out an email to *Wired*, *The Washington Post*, and other news organizations, that sought to blame the victim. "The AI chatbot Tay," the spokesperson explained, "is a machine learning project, designed for human engagement. It is as much a social and cultural experiment, as it is technical. Unfortunately, within the first 24 hours of coming online, we became aware of a coordinated effort by some users to abuse Tay's commenting skills to have Tay respond in inappropriate ways. As a result, we have taken Tay offline and are making adjustments" (Risely, 2016). According to Microsoft, it is not the programmers or the corporation who are responsible for the hate speech. It is the fault of the users (or some users) who interacted with Tay and taught her to be a bigot. Tay's racism, in other word, is our fault. Later, on 25 March 2016, Peter Lee, VP of Microsoft Research, posted the following apology on the Official Microsoft Blog: "As many of you know by now, on Wednesday we launched a chatbot called Tay. We are deeply sorry for the unintended offensive and hurtful tweets from Tay, which do not represent who we are or what we stand for, nor how we designed Tay. Tay is now offline and we'll look to bring Tay back only when we are confident we can better anticipate malicious intent that conflicts with our principles and values" (Microsoft, 2016b). But this apology is also frustratingly unsatisfying or interesting (it all depends on how you look at it). According to Lee's carefully worded explanation, Microsoft is only responsible for not *anticipating* the bad outcome; it does not take responsibility for the offensive tweets. For Lee, it is Tay who (or "that," and words matter here) is named and recognized as the source of the "wildly inappropriate and reprehensible words and images" (Microsoft, 2016b). And since Tay is a kind of "minor" (a teenage AI) under the protection of her parent corporation, Microsoft needed to step-in, apologize for their "daughter's" bad behavior, and put Tay in a time out.

Although the extent to which one might assign "agency" and "responsibility" to these mechanisms remains a contested issue, what is not debated is the fact that the rules of the game have changed significantly. As Andreas Matthias (2004) points out, summarizing his survey of learning automata:

> Presently there are machines in development or already in use which are able to decide on a course of action and to act without human intervention. The rules by which they act are not fixed during the production process, but can be changed during the operation of the machine, by the machine itself. This is what we call machine learning. Traditionally we hold either the operator/manufacture of the machine responsible for the consequences of its operation or "nobody" (in cases, where no personal fault can be identified). Now it can be shown that there is an increasing class of machine actions, where the traditional ways of responsibility ascription are not compatible with our sense of justice and the moral framework of society because nobody has enough control over the machine's actions to be able to assume responsibility for them (p. 177).

In other words, the instrumental theory of things, which had effectively tethered machine action to human agency, no longer adequately applies to mechanisms that have been deliberately designed to operate and exhibit some form, no matter how rudimentary, of independent action or autonomous decision making. Contrary to the usual instrumentalist way of thinking, we now have things that are deliberately designed to exceed our control and our ability to respond or to answer for them.

2.3 Things that are More than Things

In July of 2014 the world got its first look at Jibo. Who or what is Jibo? That is an interesting and important question. In a promotional video that was designed to raise capital investment through pre-orders, social robotics pioneer Cynthia Breazeal introduced Jibo with the following explanation: "This is your car. This is your house. This is your toothbrush. These are your things. But these [and the camera zooms into a family photograph] are the things that matter. And somewhere in between is this guy. Introducing Jibo, the world's first family robot" (Jibo 2014). Whether explicitly recognized as such or not, this promotional video leverages a crucial ontological distinction that Jacques Derrida (2005) calls the difference between "who" and "what" (p. 80). On the side of "what" we have those things that are mere instruments—our car, our house, and our toothbrush. According to the usual way of thinking, these things are mere instruments or tools that do not have any independent status whatsoever. We might worry about

the impact that the car's emissions has on the environment (or perhaps stated more precisely, on the health and well-being of the other human beings who share this planet with us), but the car itself is not a socially significant subject. On the other side there are, as the video describes it "those things that matter." These things are not things, strictly speaking, but are the other persons who count as socially and morally significant Others. Unlike the car, the house, or the toothbrush, these Others have independent status and can be benefitted or harmed by our decisions and actions.

Jibo, we are told, occupies a place that is situated somewhere in between what are mere things and those Others who really matter. Consequently Jibo is not just another instrument, like the automobile or toothbrush. But he/she/it (and the choice of pronoun is not unimportant) is also not quite another member of the family pictured in the photograph. Jibo inhabits a place in between these two ontological categories. It is a kind of "quasi-other" (Ihde, 1990, p. 107). This is, it should be noted, not unprecedented. We are already familiar with other entities that occupy a similar ambivalent social position, like the family dog. In fact animals, which since the time of Rene Descartes have been the other of the machine (Gunkel, 2012a, p. 60), provide a good precedent for understanding the changing nature of things in the face of social robots, like Jibo. "Looking at state of the art technology," Kate Darling (2012) writes, "our robots are nowhere close to the intelligence and complexity of humans or animals, nor will they reach this stage in the near future. And yet, while it seems far-fetched for a robot's legal status to differ from that of a toaster, there is already a notable difference in how we interact with certain types of robotic objects" (p. 1). This occurs, Darling continues, because of our tendencies to anthropomorphize things by projecting into them cognitive capabilities, emotions, and motivations that do not necessarily exist in the mechanism per se. But it is this emotional reaction that necessitates new forms of obligation in the face of things. "Given that many people already feel strongly about state-of-the-art social robot 'abuse,' it may soon become more widely perceived as out of line with our social values to treat robotic companions in a way that we would not treat our pets" (Darling, 2012, p. 1).

Jibo, and other social robots like it, are not science fiction. They are already or will soon be in our lives and in our homes. As Breazeal (2002) describes it, "a sociable robot is able to communicate and interact with us, understand and even relate to us, in a personal way. It should be able to understand us and itself in social terms. We, in turn, should be able to understand it in

the same social terms—to be able to relate to it and to empathize with it…In short, a sociable robot is socially intelligent in a human-like way, and interacting with it is like interacting with another person" (p. 1). In the face of these socially situated and interactive entities we are going to have to decide whether they are mere things like our car, our house, and our toothbrush; someone who matters like another member of the family; or something altogether different that is situated in between the one and the other. In whatever way this comes to be decided, however, these things will undoubtedly challenge the way we typically distinguish between who is to be considered another social subject and what remains a mere instrument or tool.

3 Between a Bot and a Hard Place

Although things are initially experienced and revealed in the mode of being Heidegger calls *Zuhandenheit* (e.g. instruments that are useful or handy for our purposes and endeavors), things do not necessarily end here. They can also, as Heidegger (1962) explains, be subsequently disclosed as present-at-hand, or *Vorhandenheit*, revealing themselves to us as objects that are or become, for one reason or another, *un-ready-to-hand* (p. 103). This occurs when things, which had been virtually invisible instruments, fail to function as they should or are designed to get in the way of their own instrumentality. "The equipmental character of things," Silvia Benso (2000) writes, "is explicitly apprehended *via negativa* when a thing reveals its unusability, or is missing, or 'stands in the way'" (p. 82). And this is what happens with things like chatterbots, machine learning applications, and social robots insofar as they interrupt or challenge the smooth functioning of their instrumentality. In fact, what we see in the face of these things is not just the failure of a particular piece of equipment—e.g. the failure of a bot like "Ivana" to successfully pass as another person in conversational interactions or the unanticipated and surprising effect of a Twitterbot like Tay that learned to be a neo-Nazi racist—but the limit of the standard instrumentalist way of thinking itself. In other words, what we see in the face chatterbots, machine learning algorithms, and social robots are things that intentionally challenge and undermine the standard way of thinking about and making sense of things. Responding to this challenge (or opportunity) leads in two apparently different and opposite directions.

3.1 Instrumentalism Redux

We can try to respond to these things as we typically have, treating these increasingly social and interactive mechanisms as mere instruments or tools. "Computer systems," as Johnson (2006) explains, "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" (p. 197). This argument is persuasive, precisely because it draws on and is underwritten by the usual understanding of things. Things—no matter how sophisticated, intelligent, and social they are, appear to be, or may become—are and will continue to be tools of human action, nothing more. If something goes wrong (or goes right) because of the actions or inactions of a bot or some other thing, there is always someone who is ultimately responsible for what happens with it. Finding that person (or persons) may require sorting through layer upon layer of technological mediation, but there is always someone—specifically some human someone—who is presumed to be responsible and accountable for it. According to this way of thinking, all things, no matter how sophisticated or interactive they appear to be, are actually "Wizard of Oz technology."[4] There is always "a man behind the curtain," pulling the strings and responsible for what happens. And this line of reasoning is entirely consistent with current legal practices. "As a tool for use by human beings," Matthew Gladden (2016) argues, "questions of legal responsibility…revolve around well-established questions of product liability for design defects (Calverley 2008, 533; Datteri 2013) on the part of its producer, professional malpractice on the part of its human operator, and, at a more generalized level, political responsibility for those legislative and licensing bodies that allowed such devices to be created and used" (p. 184).

But this strict re-application of instrumentalist thinking, for all its usefulness and apparent simplicity, neglects the social presence of these things and the effects they have within the networks of contemporary culture. We are, no doubt, the ones who design, develop, and deploy these technologies, but what happens with them once they are "released into the wild" is not necessarily predictable or completely under our control. In fact, in situations where something has gone wrong, like the Tay incident, or gone right, as was the case with AlphaGo, identifying the responsible party or parties behind these things is at least as difficult as ascertaining the "true

identity" of the "real person" behind the avatar. Consequently things like mindless chatterbots, as Mowbray (2002) points out, do not necessarily need human-level intelligence, consciousness, sentience, etc. to complicate questions regarding responsibility and social standing. Likewise, as Reeves and Nass (1996) already demonstrated over two decades ago with things that were significantly less sophisticated than these recent technological innovations, we like things. And we like things even when we know they are just things. "Computers, in the way that they communicate, instruct, and take turns interacting, are close enough to human that they encourage social responses. The encouragement necessary for such a reaction need not be much. As long as there are some behaviors that suggest a social presence, people will respond accordingly… Consequently, any medium that is close enough will get human treatment, even though people know it's foolish and even though they likely will deny it afterwards" (p. 22). For this reason, reminding users that they are just interacting with "mindless things," might be the "correct information," but doing so is often as ineffectual as telling movie-goers that the action they see on the screen is not real. We know this, but that does not necessarily change things. So what we have is a situation where our theory concerning things—a theory that has considerable history behind it and that has been determined to be as applicable to simple devices like hand tools as it is to complex technological systems—seems to be out of sync with the actual experiences we have with things in a variety of situations and circumstances. In other words, the instrumentalist way of thinking may be ontologically correct, but it is socially inept and out of touch.

3.2 Thinking Otherwise or the Relational Turn

As an alternative, we can think things otherwise. This other way of thinking effectively flips the script on the standard way of dealing with things whereby, as Luciano Floridi (2013) has describes it, what something is determines how it is treated (p. 116). Thinking otherwise deliberately inverts and distorts this procedure by making the "what" dependent on and derived from the "how." The advantage to this way of thinking is that it not only provides an entirely different method for responding to the social opportunities and challenges of all kind of things—like chatterbots, learning algorithms, and social robots—but also formulates an entirely different way of thinking about things in the face of others, and others forms of otherness. Following the contours of this alternative way of thinking, something's status—its social, moral and even

ontological situation—is decided and conferred not on the basis of some pre-determined criteria or capability (or lack thereof) but in the face of actual social relationships and interactions. "Moral consideration," as Mark Coeckelbergh (2010) describes it, "is no longer seen as being 'intrinsic' to the entity: instead it is seen as something that is 'extrinsic': it is attributed to entities within social relations and within a social context" (p. 214). In other words, as we encounter and interact with others—whether they be other human persons, other kinds of living beings like animals or plants, the natural environment, or a socially interactive bot—this other entity is first and foremost situated in relationship to us. Consequently, the question of something's status does not necessarily depend on what it is in its essence but on how she/he/it (and the pronoun that comes to be deployed in this circumstance is not immaterial) supervenes before us and how we decide to respond (or not) "in the face of the other," to use terminology borrowed from Emmanuel Levinas (1969). In this transaction, "relations are prior to the things related" (Callicott, 1989, p. 110), instituting what Anne Gerdes (2015), following Coeckelbergh (2012, p. 49) and myself (Gunkel, 2012), has called "the relational turn."

This shift in perspective, it is important to point out, is not just a theoretical game, it has been confirmed in numerous experimental trials and practical experiences with things. The computer as social actor (CASA) studies undertaken by Reeves and Nass (1996), for example, demonstrated that human users will accord computers social standing similar to that of another human person and this occurs as a product of the extrinsic social interaction, irrespective of the actual composition (or "being" as Heidegger would say) of the thing in question. These results, which were obtained in numerous empirical studies with human subjects, have been independently verified in two recent experiments with robots, one reported in the *International Journal of Social Robotics* (Rosenthal-von der Pütten et al, 2013), where researchers found that human subjects respond emotionally to robots and express empathic concern for machines irrespective of knowledge concerning the actual ontological status of the mechanism, and another that used physiological evidence, documented by electroencephalography, of the ability of humans to empathize with what appears to be "robot pain" (Suzuki et al, 2015). And it appears that this happens not just with seemingly intelligent artifacts in the laboratory setting but with just about any old thing that has some social presence, like the very industrial-looking Packbots that are being utilized on the battlefield. As P. W. Singer (2009, p. 338) has reported, soldiers form surprisingly close personal bonds with their units' Packbot, giving them names,

awarding them battlefield promotions, risking their own lives to protect that of the machine, and even mourning their "death." This happens, Singer explains, as a product of the way the mechanism is situated within the unit and the social role that it plays in field operations. And it happens in direct opposition to what otherwise sounds like good common sense: They are just things—instruments or tools that feel nothing.

Once again, this decision sounds reasonable and justified. It extends consideration to these other socially aware and interactive things and recognizes, following the predictions of Wiener (1950, p. 16), that the social situations of the future will involve not just human-to-human interactions but relationships between humans and machines and machines and machines. But this shift in perspective also has significant costs. For all its opportunities, this approach is inevitably and unavoidably exposed to the charge of relativism—"the claim that no universally valid beliefs or values exist" (Ess, 1996, p. 204). To put it rather bluntly, if the social status of things is relational and open to social negotiation, are we not at risk of affirming a kind of social constructivism or 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 (2000) strategy of "fully endorsing what one is accused of" (p. 3). So yes, relativism, but an extreme and carefully articulated version of it. That is, a relativism (or, if you prefer, a "relationalism") that can no longer be comprehended by that kind of understanding of the term which makes it the mere negative and opposite of an already privileged universalism. Relativism, therefore, does not necessarily need to be construed negatively and decried, as Žižek (2006) himself has often done, as the epitome of postmodern multiculturalism run amok (p. 281). It can be understood otherwise. "Relativism," as Robert Scott (1976) argues, "supposedly, means a standardless society, or at least a maze of differing standards…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" (p. 264). 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." Instead, what is obtained is a kind of socially attentive thinking that turns out to be much more responsive and responsible in the face of other things.

These two options anchor opposing ends of a spectrum that can be called *the machine question* (Gunkel 2012a). How we decide to respond to the opportunities and challenges of this 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 socially significant subjects, and what things 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. We are, therefore, responsible both for deciding who counts as another subject and what is not and, in the process, for determining the way we perceive the current state and future possibility of social relations.

Notes

[1] A consequence of this way of thinking about things is that all things are initially revealed and characterized as *media* or something *through* which human users act. For more on this subject, see *Heidegger and the Media* (Gunkel and Taylor, 2014).

[2] Identification of these two alternatives have also been advanced in the phenomenology of technology developed by Don Ihde. In *Technology and the Lifeworld*, Ihde (1990) distinguishes between "those technologies that I can take into my experience that through their semi-transparency they allow the world to be made immediate" and "alterity relations in which the technology becomes quasi-other, or technology "as" other *to* which I relate" (p. 107).

[3] Although the term "chatterbot" was not utilized by Weizenbaum, it has been applied retroactively as a result of the efforts of Michael Mauldin, founder and chief scientist of Lycos, who introduced the neologism in 1994 in order to identify a similar NLP application that he eventually called Julia.

[4] "Wizard of Oz" is a term that is utilized in Human Computer Interaction (HCI) studies to describe experimental procedures where test subjects interact with a computer system or robot that is assumed to be autonomous but is actually controlled by an experimenter who remains hidden from view. The term was initially introduced by John F. Kelly in the early 1980s.

References

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

Bessi, A. and E. Ferrara (2016). Social Bots Distort the 2016 U.S. Presidential Election Online Discussion. *First Monday* 21(11). http://firstmonday.org/ojs/index.php/fm/article/view/7090/5653

Breazeal, C. L. (2004). *Designing Sociable Robots*. Cambridge, MA: MIT Press.

Callicott, J. B. (1989). *In Defense of the Land Ethic: Essays in Environmental Philosophy*. Albany, NY: State University of New York Press.

Calverley, D. J. (2008). Imaging a Non-Biological Machine as a Legal Person. *AI & Society* 22(4): 523-537.

Coeckelbergh, M. (2010). Robot Rights? Towards a Social-Relational Justification of Moral Consideration. *Ethics and Information Technology* 12: 209-221.

Coeckelbergh, M. (2012). Growing Moral Relations: A Critque of Moral Status Ascription. New York: Palgrave Macmillan.

Darling, K. (2012). Extending Legal Protection to Social Robots. *IEEE Spectrum*, 10 September 2012. http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/extending-legal-protection-to-social-robots

Datteri, E. (2013). Predicting the Long-Term Effects of Human-Robot Interaction: A Reflection on Responsibility in Medical Robotics. *Science and Engineering Ethics* 19(1): 139-160, 2013.

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

Epstein, R. (2007). From Russia, With Love: How I Got Fooled (And Somewhat Humiliated) by a Computer. *Scientific American Mind* Oct/Nov: 16-17.

Ess, C. (1996). The Political Computer: Democracy, CMC, and Habermas. In C. Ess (ed.), *Philosophical Perspectives on Computer-Mediated Communication* (pp. 197-232). SUNY Press, Albany, NY,

Feenberg, A. (1991). *Critical Theory of Technology.* New York: Oxford University Press.

Floridi, L. (2013). *The Ethics of Information*. Oxford: Oxford University Press.

Gartner (2013). Press Release. http://www.gartner.com/newsroom/id/2636073

Gerdes, A. (2015). The Issue of Moral Consideration in Robot Ethics. *ACM SIGCAS Computers & Society* 45(3): 274-280.

Gladden, M. E. (2016) The Diffuse Intelligent Other: An Ontology of Nonlocalizable Robots as Moral and Legal Actors. In M. Nørskov, editor, *Social Robots: Boundaries, Potential, Challenges*, pp. 177-198. Burlington, VT: Ashgate.

Go Ratings (20 November 2016). https://www.goratings.org/

Google DeepMind (2016). AlphaGo. https://deepmind.com/alpha-go.html

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

Gunkel, D. J. (2012b). Communication and Artificial Intelligence: Opportunities and Challenges for the 21st Century. *Communication +1* 1(1): 1-25. http://scholarworks.umass.edu/cpo/vol1/iss1/1/

Gunkel, D. J. and P. A. Taylor (2014). *Heidegger and the Media*. Cambridge: Polity.

Hall, J. S. 2001. Ethics for Machines. *KurzweilAI.net*. http://www.kurzweilai.net/ethics-for-machines

Harman, G. (2002). *Tool Being: Heidegger and the Metaphysics of Objects*. Peru, IL: Open Court Publishing.

Heidegger, M. (1962). *Being and Time*. Trans. by J. Macquarrie and E. Robinson. New York: Harper and Row.

Heidegger, M. (1977). *The Question Concerning Technology and Other Essays*. Trans. by William Lovitt New York: Harper & Row.

Hingston, P. (2014). *Believable Bots: Can Computers Play Like People?* New York: Springer.

Ihde, D. (1990). *Technology and the Lifeworld: From Garden to Earth*. Bloomington, IN: Indiana University Press.

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

Johnson, D. G. (1985). *Computer Ethics*. Upper Saddle River, NJ: Prentice Hall.

Johnson, D. G. (2006). Computer Systems: Moral Entities but not Moral Agents. *Ethics and Information Technology* 8: 195-204.

Levinas, E. (1969). *Totality and Infinity: An Essay on Exteriority.* Trans. by Alphonso Lingis. Pittsburgh, PA: Duquesne University.

Matthias, A. (2004). The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata. *Ethics and Information Technology* 6: 175-183.

McLuhan, M. and Q. Fiore (2001). *War and Peace in the Global Village.* Berkeley, CA: Ginko Press.

Metz, C. (2016). Google's AI Wins a Pivotal Second Game in Match with Go Grandmaster. *Wired*. http://www.wired.com/2016/03/googles-ai-wins-pivotal-game-two-match-go-grandmaster/

Microsoft. (2016a). Meet Tay—Microsoft A.I. Chatbot with Zero Chill. https://www.tay.ai/

Microsoft. (2016b). Learning from Tay's introduction. *Official Microsoft Blog*. https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/

Mowbray, M. (2002). Ethics for Bots. Paper presented at the 14th International Conference on System Research, Informatics and Cybernetics. Baden-Baden, Germany. 29 July-3 August. http://www.hpl.hp.com/techreports/2002/HPL-2002-48R1.pdf

Peterson, A. (13 August 2013). On the Internet, No one Knows You're a Bot. And That's a Problem. *The Washington Post*. https://www.washingtonpost.com/news/the-switch/wp/2013/08/13/on-the-internet-no-one-knows-youre-a-bot-and-thats-a-problem/?utm_term=.b4e0dd77428a

Reeves, B. and C. Nass. (1996). The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge: Cambridge University Press.

Risely, J. (2016). Microsoft's Millennial Chatbot Tay.ai Pulled Offline After Internet Teaches Her Racism. *GeekWire*. http://www.geekwire.com/2016/even-robot-teens-impressionable-microsofts-tay-ai-pulled-internet-teaches-racism/

Rosenthal-von der Pütten, A. M., N. C. Krämer, L. Hoffmann, S. Sobieraj and S. C. Eimler (2013). An Experimental Study on Emotional Reactions Towards a Robot. *International Journal of Social Robotics* 5: 17-34.

Scott, R. L. (1976). On Viewing Rhetoric as Epistemic: Ten Years Later. *Central States Speech Journal* 27(4): 258-266.

Singer, P. W. (2009). *Wired for War: The Robotics Revolution and Conflict in the Twenty-First Century*. New York: Penguin Books.

Suzuki, Y., L. Galli, A. Ikeda, S. Itakura and M. Kitazaki (2015). Measuring Empathy for Human and Robot Hand Pain Using Electroencephalography. *Scientific Reports* 5: 15924. http://www.nature.com/articles/srep15924

Tanz, J. (2016). The End of Code. *Wired* 24(6): 75-79. http://www.wired.com/2016/05/the-end-of-code/

Turing, A. (1999). Computing Machinery and Intelligence. In P. A. Meyer (ed.), *Computer Media and Communication: A Reader*, 37-58. Oxford: Oxford University Press.

Turing, A. (2004). Can Automatic Calculating Machines Be Said to Think? *The Essential Turing*, 487-505. Oxford: Oxford University Press.

Weizenbaum, J. (1976). *Computer Power and Human Reason: From Judgment to Calculation*. San Francisco, CA: W. H. Freeman and Company.

Wiener, N. (1950). *The Human Use of Human Beings: Cybernetics and Society*. Boston: Ad Capo Press.

Zeifman, I. (2017). Bot Traffic Report 2016. *Incapsula*. https://www.incapsula.com/blog/bot-traffic-report-2016.html

Zimmerman, M. E. (1990). *Heidegger's Confrontation with Modernity: Technology, Politics, Art*. Bloomington, IN: Indiana University Press.

Žižek, S. (2000). *The Fragile Absolute or, Why Is the Christian Legacy Worth Fighting For?* New York: Verso.

Žižek, S. (2006). *The Parallax View*. Cambridge, MA: MIT Press.