Nudging for Good: Robots and the Ethical Appropriateness of Nurturing Empathy and Charitable Behavior

Jason Borenstein* and Ron Arkin**

Predictions are being commonly voiced about how robots are going to become an increasingly prominent feature of our day-to-day lives. Beyond the military and industrial sectors, they are in the process of being created to function as butlers, nannies, housekeepers, and even as companions (Wallach and Allen 2009). The design of these robotic technologies and their use in these roles raises numerous ethical issues. Indeed, entire conferences and books are now devoted to the subject (Lin et al. 2014).[1] One particular under-examined aspect of human-robot interaction that requires systematic analysis is whether to allow robots to influence a user's behavior for that person's own good. However, an even more controversial practice is on the horizon and warrants attention, which is the ethical acceptability of allowing a robot to "nudge" a user's behavior for the good of society.

For the purposes of this paper, we will examine the feasibility of creating robots that would seek to nurture a user's empathy towards other human beings. We specifically draw attention to whether it would be ethically appropriate for roboticists to pursue this type of design pathway. In our prior work, we examined the ethical aspects of encoding Rawls' Theory of Social Justice into robots in order to encourage users to act more socially just towards other humans (Borenstein and Arkin 2016). Here, we primarily limit the focus to the performance of charitable acts, which could shed light on a range of socially just actions that a robot could potentially elicit from a user and what the associated ethical concerns may be. In short, should robots be deliberately designed to subtly and/or overtly nudge a user's behavior in the hopes that this strategy may promote the good of society? Who determines what is good for society in this context? Are there any universal social goods that should be considered? What role, if any, do cultural variations and tolerances have in this context? The dystopian use of technology for "social good" was famously framed by Orwell's *1984*. Along these lines, we believe that the practice of nudging user behavior deserves greater attention from professional communities given how pervasive robots may become and how embedded nudging strategies already are in technological devices.

1. Robots and Nudges

A "nudge", as defined by Thaler and Sunstein (2008), is as an attempt to mold or guide behavior without relying on legal or regulatory mechanisms. Many of the technologies that influence user behavior are relational artifacts (Turkle 2005), but a user does not necessarily have to feel they have some form of reciprocal relationship with an object in order for it to nudge the person's behavior (for example, a speed bump causing a driver to slow down). There is also overlap between the concepts of persuasive technology and nudging but a person can certainly be nudged without the use of technology. A customer could be "nudged" towards the selection of a particular product through the use of scents, sounds, or images or by eliciting certain attitudes or beliefs from the customer (e.g., this product will make you "more attractive").

President Obama's administration has made extensive use of insights from behavioral economics to nudge American citizens towards preferred courses of action, such as being more honest on tax returns (White House 2015). For example, the strategy of informing one taxpayer on a tax form that other citizens have paid their taxes accurately and on time can exert peer pressure on that person to be in compliance (Vinik 2015). State agencies are also experimenting with nudges, including by developing programs for a prisoner's "own good" as it pertains to paying child support (Bohannon 2016).

Similar to other computing technology (such as fitness trackers or health-related apps), robots have much potential to nudge user behavior. Given that robots are and will be built in many shapes and sizes, they can interact with users in a broad range of ways, from the instructive to the intimate. But the focus here is on physically-embodied robots that could serve as friends, caregivers, or assistants for human beings. These companion robots could take

* Director of Graduate Research Ethics Programs and Associate Director of the Center for Ethics and Technology, School of Public Policy and Office of Graduate Studies, Georgia Institute of Technology
** Regents' Professor, School of Interactive Computing within the College of Computing, Georgia Institute of Technology

[1] For an example of a conference series, see http://conferences.au.dk/robo-philosophy/ Accessed 26 June 2016.

many forms; for example, Paro, Jibo, Domgy, Nao, and Pepper can fall within this category.[2] While the ability of robots to have complex, sustained interactions with humans over prolonged durations is currently lacking, it may be on the horizon. Roboticists envision a time when their technology will become more sophisticated than it is today by orders of magnitude; a robot that is available to interact 24/7 may await in the future.

What makes robots distinctive as compared to other computing technology is their potential to go beyond the realm of merely verbal and/or two-dimensional visual expressions. Spatial proximity (proxemics), body language (kinesics), and touch are facets of a physically-embodied robot that would enable it to interact in ways that a mobile phone, tablet, or other standard computing devices cannot match (Brooks and Arkin 2007). A robot, like *Nao* or *Pepper* for example, could gesture towards an object in the physical environment, which could indicate to a user that the object should be picked up and moved out of the way. Or, the robot could pat someone on the back in order to provide comfort. A robot's interactive possibilities offer distinct advantages in terms of how it could nudge a user's behavior.

2. Will Robots Be Effective At Nudging Users?

The subtlety of nudges combined with the power that they can exert over a target individual's decision-making process makes them a noteworthy source of ethical concern. Due to its physical presence, a robot companion would certainly have various opportunities to guide a user towards a certain behavioral direction (Borenstein and Arkin 2016). Yet it is an open question regarding how effective companion robots will be at influencing user behavior, and more specifically, whether they have the capacity to nurture human empathy. Indeed, research is already proceeding in this area (Pettinati and Arkin 2015; Zuckerman and Hoffman 2015). Although we will not fully address the associated complexities here, we will briefly outline some of relevant dimensions below.

Technological prediction is often a perilous endeavor; it was only several decades ago, for example, that computing companies expressed doubts about the desirability of owning a personal computer; also relatively few individuals anticipated the emergence of the Internet (Nye 2004). But to focus more specifically, will the sophistication of robots proceed along predicted trajectories whereby they will at some point become an integral part of the daily lives of human beings? Current companion robots, such as Paro, are not typically used for prolonged durations and are fairly limited in terms of the intellectual engagement they can provide. A companion robot may (temporarily) be able to brighten one's mood or perhaps mitigate loneliness (e.g., see Banks et al 2008). Furthermore, the IBM cognitive system *Watson*, chatbots, and other computing technology are improving to such an extent that some argue the Turing Test has been satisfied (Hornyak 2014). Yet the conversational limitations of these entities usually become readily apparent in a relatively short amount of time; they are typically unable to participate in sustained conversations. Even one of the most sophisticated chatbots ever created, Tay, had to be removed from the Internet when it was manipulated by its human counterparts and started to utter offensive comments (Masunaga 2016). Given this state of affairs and other relevant factors, it is unlikely that a current iteration of companion robot would have a lasting effect on its human counterpart's personality or character.

Furthermore, for our purposes here, we seek to analyze whether robots may be able to nurture empathy within a user (later on, we will explore the ethical dimensions of pursuing this type of design pathway). To begin, will robots have the necessary sophistication to recognize the limits of when and how often to encourage "positive" or praiseworthy acts such as contributing to charity? If a fundamental change in the user's nature is an essential prerequisite, then perhaps not. But on the other hand, advertisers are incredibly effective at influencing the behavior of customers and seemingly this does not necessitate that the individual customer has become a "different" person. And as previously mentioned, roboticists are already in the process of developing strategies that could enable a robot to elicit empathy from a user (e.g., Hoffman et al. 2015).

According to Kosner (2014), "Habit-formation is no longer a nice-to-have but a need-to-have aspect of making a successful product." More specific to the realm of computing, coders exhibit enormous proficiency at modifying the behavior of others, including by separating people from their money (through in-app purchases, targeted advertising, etc.). Researchers, including Reeves and Nass (1996), have shown that humans interact with computing technology

[2] For images of these robots, see http://www.parorobots.com/index.asp; https://www.jibo.com/; http://www.marketwired.com/press-release/roobo-offers-sneak-peek-into-worlds-first-intelligent-pet-robot-pepcom-digital-experience-2135140.htm; https://www.ald.softbankrobotics.com/en/cool-robots/nao; https://www.ald.softbankrobotics.com/en/cool-robots/pepper. Accessed 26 June 2016.

in similar ways to how they interact with other people. And improvements in the realm of artificial intelligence may lead users to become increasingly emotionally attached to computing technology (Weigel 2016). For some time, roboticists and others have been seeking to make use of psychological studies in order to elicit desired emotional responses from users (e.g., Arkin et al 2003). For example, Ham and Midden (2013) discuss how receiving feedback from a robot might encourage users to save energy; they contend that "social" feedback is more effective at altering user behavior than feedback which is "factual" in nature. Thus if the aim is to nudge a user in the sense of causing short-term behavioral changes, then a fairly compelling case can be made that roboticists will achieve success (if permitted to do so).

3. Empathy, Charity, and the Role of Robots

Many proposed definitions of empathy are in the literature (Coplan and Goldie 2011). Yet our primary focus centers on what is often referred to as "affective empathy", which roughly refers to having the ability to view the world from someone else's perspective and sharing in that person's emotional state (Stueber 2014). Presumably, the world would be a much better place if people had more empathy for one another. Ideally, parents will dedicate significant effort towards nurturing empathy within their children. Empathy can, for example, play an important role in a child's willingness to participate in community service activities (Swick 2001). Cultivating empathy in children and others can be an important mechanism for preventing the performance of violent acts (Swick 2005). Empathy can support human dignity and is arguably an essential part of the fabric that holds a society together (Zinn 1993). While critics express doubts about whether empathy is a requisite component of morality (e.g., Prinz 2011), for the purposes of the following discussion, we assume that an important connection exists between the two.

According to Swick (2005, 54), key components of empathy are being able to detect cues, including non-verbal ones, from others and having "self-other relational awareness", which entails recognizing one's own feelings and connecting them to the feelings of others. Interestingly, Konrath, O'Brien, and Hsing (2011) performed a meta-analysis of studies conducted between 1979 and 2009, and emerged with the disconcerting finding that there has been a sharp drop in empathy among college-age students during that time. Moreover, Dolby (2014, 42) suggests that without concerted efforts to cultivate empathy in forthcoming generations, human beings might not be able to handle the challenges associated with resource scarcity and climate change.

If robots could nurture empathy within users, this would arguably be a beneficial outcome. Granted that other computing devices could be used to accomplish similar goals, but to reiterate, a physical-embodied robot has advantages in that it could express itself in a range of verbal and non-verbal ways. As Zuckerman and Hoffman (2005) state, robots could seek to "increase people's self-awareness to the emotional state of others, leading to behavior change." For instance, a robot could attempt to temper sexist, racist, or homophobic tendencies by providing negative verbal feedback and altering its posture when slurs are uttered by a user. More generally, prompts that remind the user of another person's needs and suffering may have the outgrowth of establishing an emotional connection between them.

Of course, this type of design strategy is not free of ethical concerns; in fact, many non-trivial problems could result. Yet for the time being, we seek to determine whether there are plausible reasons to believe that a robot could successfully accomplish this type of goal before proceeding to examine the ethical implications of doing so. To achieve some level of success the robot would, as a starting point, have to be able to evoke emotional attachments to others in the world in a way that would resonate with the user. While establishing attachment to robots themselves has been shown to be easy to accomplish (Reeves and Nass 1996; Turkle 2011), relatively little research on a robotic device's ability to foster human-human attachment has been undertaken to date (e.g., Zuckerman and Hoffman 2015). And the research in this realm often involves the use of a robot in care settings (Chang and Sabanovic 2015; Scasselati et al. 2012).

A possible outgrowth of feeling empathy for others is the willingness to perform charitable acts. At times, being charitable could serve as a proxy for indicating that the donor is experiencing empathy for another person. Yet to our knowledge, there is no necessary causal relationship between empathy and charity. Obviously, a person does not have to be motivated by empathy in order to give to charity; in fact, guilt, economics (e.g., receiving a tax break), or quid pro quo, might be the driving reason for doing so. However, a hoped for by-product of efforts to nurture empathy would be the increasing performance of such acts. Even as we are envisioning how robots might encourage users to perform these acts, we again do not assume that the technology, at least in the near term, will be able to fundamentally change a user's character or personality.

In the future, the sophistication of robots may advance so much that they become integrated into a person's daily life and perhaps even intimate relationships may form (Levy 2007; Whitby 2012). At some point, a user's psychological health and well-being may be acutely intertwined with the technology. Yet for now, it is far more likely the case that robots would be successful at altering a user's attitude or mood (Moshkina and Arkin 2005). Thus, our primary aim is to consider the potential that companion robots have to foster empathy in a user with the goal of nudging the user towards performing charitable acts.

4. A Robot's Capacity for Nurturing Empathy: Reasons for Skepticism

A persistent concern emerging out of the digital age is whether computing technology may be eroding the empathy that individuals have for one another (Turkle 2011). Vicious and brutal cyber bullying may be illustrative of this point. For example, a mother used to a pseudonym to create an online account in order to communicate with her daughter's former friend; the resulting online exchanges allegedly contributed to the former friend's suicide (ABC News 2007). Women who play online games have at times been subjected to threats of physical violence and rape (Mettler 2016). Rude behavior in Internet forums is so pervasive that many companies rely on users to flag offensive content for removal. Furthermore, the European Union has requested that several social media companies develop strategies for curtailing the hate speech of users on their websites (Hern 2016).

For some time now, scholars have warned that "risk-free" online behavior is having deleterious effects on the psychological and moral development of Internet users (see e.g., Dreyfus 2004). Along these lines, Turkle (2011) has been a frequent critic of computing technology in large part because of its potential to lessen the frequency and quality of human-human interaction; the overuse of the technology could arguably stifle the development of valuable social skills and lead to a decrease in empathy for other people. In one of her recent books (2015), she contends that mobile phones are undermining the capacity for having meaningful conversations.

The availability of more advanced robots will lead to diminished human-human contact for at least some segments of society. And if there is good reason to believe that interacting with other humans is a necessary condition for acquiring empathy, then this could be a profoundly troubling occurrence. Moreover, if a user perceives bad behavior towards a robot as being "consequence-free", one could certainly be skeptical about whether the user would develop sufficient empathy towards other humans. Quite the contrary, a user's anti-social behaviors could become entrenched without the appropriate social cues and feedback to deter the performance of such behaviors.

Another reason for skepticism is that present-day robots can only display "empathic behavior" and not genuine empathy (Stephan 2015, 111). Correspondingly, at least some users may perceive a robot's attempt to appear empathic, with the goal of nurturing empathy in the user, as being condescending, disrespectful, or otherwise inappropriate. Furthermore, Stephan argues that "empathy—understood as a genuine human capacity—is not task specific" (2015, 114). In other words, if a robot only seems to appear empathic in limited circumstances, for example when a user utters certain key words, then it could undermine the effectiveness of the attempt to foster empathy in the user.

5. A Robot's Capacity for Nurturing Empathy: Reasons for Optimism

There are, however, many counterbalancing reasons to believe that robots could foster empathy in human beings. For one, humans often bond with non-human entities, and positive traits can allegedly result from those interactions. Fiction writers are profoundly adept at sparking emotional responses from audiences; writers can cause readers to be emotionally invested in a fictional character in part by encouraging us to identify with the character's circumstances (Carroll 2011). Moreover, having a pet arguably contributes to the development of empathy in children (Endenburg and van Lith 2011).

The extensive knowledge that roboticists collectively possess in the realm of affective computing and the insights that they glean from studies of how empathy emerges can provide significant in-roads for creating conditions that mimic how empathy "naturally" forms between human beings. In addition, a robot's design offers avenues for complex, relevant feedback; for example, if a user curses, a sophisticated companion robot could express disappointment through a combination of both verbal and non-verbal cues. Arkin's laboratory has designed an ethical governor for this specific purpose, where as specific set of rules are defined to defuse a potentially volatile situation between two human participants (Shim and Arkin 2015). Arkin (2014b) has also been developing ways for a robot to uphold the dignity of Parkinson's patients when they interact with caregivers. In this case, the robot

maintains a model of empathy of the human caregiver and monitors for a lack of congruence regarding the patient's shame and embarrassment levels and how they are perceived by a clinician. If there is a significant disparity, the robot then undertakes an action to assist in restoring a more congruent relationship between the two humans.

6. Encouraging Charitable Acts: "The How"

One indication of a robot's success at nurturing empathy is whether the user performs (more) charitable acts. Yet to reiterate, a charitable act could take place without the person performing the act feeling empathy towards others. Examples of how robots could behave include verbal prompts to donate money. If a robot possesses the capacity to construct a detailed user profile, then offering specifically tailored recommendations becomes a distinct possibility. For example, if the user has a family history of a particular illness (e.g., heart disease), the robot could suggest contributing to an associated charity (the American Heart Association). As stated by Ferenbok and colleagues, "algorithms are being taught to look ahead, to anticipate our potential actions in what is now called *predictive analytics*" (2016, 98). Presumably, this technique will be integrated into many types of companion robots.

As stated previously, physically-embodied robots can interact with a user in many ways beyond the merely verbal, and thus, it opens up various possibilities for nudge the user's behavior. Animators have known for quite some time that the size and expressiveness of their creations can elicit emotional responses from movie viewers (Thomas and Johnston 1981). Similarly, a robot's "sad eyes" or body posture (kinesics) could convey disappoint if the user does not make a donation, which could then change to a "happy" expression or mannerisms when it does occur.[3] Furthermore, positive feedback for volunteerism could encourage repeated charitable acts. Roboticists could draw from strategies used by charitable and other organizations for enhancing effective fundraising.

7. Encouraging Charitable Acts: "The When"

The timing of prompts is integrally tied into whether a nudge would be effective. These prompts would have to be tailored to the individual preferences, tolerances, and disposition of a user. For one, the robot must not be perceived as being annoying; the unsolicited call at dinner time from a telemarketer is illustrative here given how it can sometimes infuriate people. For one, the user would have to be in the "right frame of mind" to be receptive to the information from the robot. Moreover, the robot would have to be discerning enough to know when a relevant opportunity arises. This requires, at a minimum, partner modeling or a partial theory of mind (Arkin 2014a). The robot would need to be able to pick up on social cues, including when it is appropriate to talk (and not interrupt the user). If the robot can detect a "vulnerable moment" to deliver a message, a user's charitable act becomes more likely.

According to Fogg, "Timing involves many elements in the environment (ranging from the physical setting to the social context) as well as the transient disposition of the person being persuaded (such as mood, feelings of self-worth, and feelings of connectedness to others)" (2003, 43). A key variable is how much of the day the robot is within proximity of, or more generally has the ability to interact with, the user. Time of day is also important: the modeling of circadian rhythms of the user may play into this (Park, Moshkina, and Arkin 2010). Case-based reasoning (Moshkina 2011), a form of machine learning, can also draw on past experiences regarding both the when and to whom a charitable nudge should be oriented towards.

Advertisers in various media platforms have been honing their craft for many decades now. For example, they have constructed strategies for pairing up certain types of products with the target demographic of a television show airing at a specific time slot. But in the future, robots offer far more opportunity time-wise than either television or print media afford, if their potential for becoming persistent companions is truly actualized. A mobile robot's ability to follow the user, for instance, introduces possibilities to influence the user's behavior in various settings and situations. This could include while the user is watching television, eating a meal, or exercising.

8. The Ethical Appropriateness of the Design Strategy

The overarching question that we have sidestepped thus far is should the robotics community be pursuing this type of goal in the first place? Fogg points out that "If you examine the history of computing technologies, you find that many high-tech products have changed the way people think, feel, and act. But most of these changes were not planned persuasive effects of the technology; they were side effects" (2003, 17). One should certainly have ethical

[3] For example, refer to the different emotional expressions that the robot Kismet can make: http://www.ai.mit.edu/projects/humanoid-robotics-group/kismet/kismet.html. Accessed 20 May 2016.

qualms about a technology that has unintended harmful consequences for users and others. Yet the focus here is on the calculated efforts that roboticists would intentionally put forward to alter user behavior. Even if the goal is to produce good consequences for the user or other people, that design pathway can have deep ethical ramifications, some of which we discussed previously (Borenstein and Arkin 2016).

Ham and Spahn (2015, 66) highlight two levels of analysis needed when determining whether a nudge or other persuasive tactic is ethically appropriate: first, whether a certain type of pervasive technique is ethically acceptable in general and second, whether it would be appropriate for robots or other computing technology to make use of that technique. For example, seemingly it is ethically appropriate to nudge people to not smoke by putting harsh, and sometimes graphic, warning labels on cigarettes. But would it be appropriate for a robot to utter similar sentiments while a person is deciding whether to smoke in a public place, especially if it is legal to do so? The robot's nudge in this case would not only be intended to benefit the individual's health, but it could protect the health of others as well. Yet Ham and Spahn indicate a robot may be a particular source of ethical concern because of the asymmetry between it and the user (i.e., presumably a robot could persuade the user but not vice versa) (2015, 66).

A deliberate nudge from a technological artifact could obviously intrude on the autonomy of a user and display a lack of respect for persons. Yet this concern alone does not unilaterally resolve whether the tactic is ethically appropriate. Many nuances associated with a "nudging" design pathway warrant thorough exploration; reflection and analysis should inform decisions about whether roboticists and others should be permitted to pursue the strategy. Our goal here is to at least generate a starting point for a conversation about the topic. With that in mind, factors that could have a direct bearing on whether a nudge is ethically acceptable include:

- Is the person consciously aware of the nudge, and if so, to what degree?
- Does the nudge engage "rational" forms of thinking?
- Can the person exert any control over whether and how the nudge occurs?
- Is the person able to avoid the nudge if that was preferred?
- Is the act that is supposed to result from the nudge likely to be beneficial to the person?
- Would the person who is the target of the nudge perceive the act as being beneficial?
- Is the intended primary beneficiary someone other than the person who is the target of the nudge?
- Are negative consequences likely to follow for failing to act in accordance with the nudge?
- Could the nudge potentially contribute to the formation of habitual behavior, or even addiction, in the person?
- Is the person particularly vulnerable to manipulation due to age, physical or mental characteristics, socio-economic status, or other relevant factors?

A key facet of the discussion is whether and how information about the robot's nudge will be conveyed to a user. Hausman and Welch suggest that "there may be something more insidious about shaping choices than about open constraint" (2010, 130). To clarify this point, they describe the relatively problematic nature of subliminal messages as compared to a law that requires drivers to wear seat belts. The former may be more intrusive and "a greater threat to liberty" because the recipient is probably unaware that it is taking place (Hausman and Welch 2010, 131). If a nudge were to constitute a "non-rational" form of persuasion, then they believe it would be ethically problematic (Hausman and Welch 2010, 135). Yet even if a robot "transparently" attempts to influence a user, the nudge may still be an effective tactic regardless. In short, is it ethically acceptable to manipulate a person's behavior even when the target person is aware the manipulation is taking place?

According to Eyal and Hoover (2013), companies are keenly aware of how to design their products in such a way to form habits in the person who interacts with them. Roboticists could build on this knowledge base and take advantage of well-established habit-forming strategies. Yet the ethical appropriateness of a nudge that contributes to habit formation can certainly be questioned, especially if some type of addiction emerges. It could even turn out to be problematic in the case of charitable contributions if, for example, users start donating so much money that the practice becomes detrimental to them (e.g., they no longer have enough money to pay their bills) or if they fail to save enough to uphold their familial obligations (e.g., to save for their children's education). An issue underlying these complexities is to whom the charitable acts should be directed. One would think that robot should not exert "too much" control over that decision; instead, the user should retain decision-making power. Yet how should a robot be advising, if at all, users to balance their needs against those of family members, local groups, and others?

According to Harris (2016), "Everyone innately responds to social approval, but some demographics (teenagers) are more vulnerable to it than others. That's why it's so important to recognize how powerful designers are when they exploit this vulnerability." Along these lines, children, those with certain types of mental impairments, lonely individuals, older adults, and others may be particularly susceptible to being manipulated by computing technology (Foog 2003, 230-232). In some situations, companion robots may end up interacting with vulnerable populations. In fact, some robots, including *Keepon*, were originally designed to interact with special needs children, and *Paro* is a robot specifically designed to provide companionship for older adults.

Moreover, a fundamental issue interwoven into this discussion is the prudence of (technological) efforts to nurture empathy in human beings. There could be critiques of the entire enterprise. For instance, Prinz questions whether empathy is a suitable basis for morality in part because he argues that it may "lead to preferential treatment" and "be subject to unfortunate biases" (2011, 226). Someone could, for example, begin to form empathy for a radical group that has an antisocial agenda. Arguably, having greater empathy for others does not invariably lead to ethical behavior (Bloom 2015). Yet we do not seek to resolve philosophical disputes about the nature and value of empathy; rather, our point is to illustrate the overarching difficulty of identifying a specific, suitable, and operationalizable objective for roboticists to pursue even if consensus could be reached that the general aim of companion robots is to encourage users to become "better humans".

9. Other Related Ethical Concerns

Given that the specific context here is on a robot's capacity to encourage a user to perform behaviors that benefit other people, another essential question to address is who has the moral authority to determine what is a worthwhile social cause? And, if roboticists become efficient at influencing user behavior, who are the appropriate individuals to wield this kind of power? Many would argue, in a similar vein to why there are ethical objections to subliminal messages, that it would be an exploitation of a user's psychological make-up. Moreover, could hackers take advantage of the persuasive techniques roboticists develop to have robots redirect users towards ethically dubious goals? But are these concerns tempered by the potential benefits to society as a whole? And who should make that decision?

Along related lines, one can envision how gathered information about an individual could be taken advantage of for other purposes (e.g., to sell products to a user). For example, the *Hello Barbie* doll can listen to children talking, send the data to cloud-based software, and then offer relevant replies to the conversation; as expected, parents complain that this could enable companies to covertly advertise to their children (Alba 2015). Privacy is a fundamental concern that emerges here. But placing that aside, the main point we are trying to convey here is that once a robot builds a profile about the user, countless exploitative possibilities arise.

Moreover, roboticists need to consider the ramifications of the ELIZA effect as it pertains to this context (Turkle 1995, 101). In short, the user may project characteristics onto a project than it does not truly have and believe a robot is capable of far more than it actually is. For instance, users may think that a robot is happy if it smiles after a charitable donation has been made. At the present time, a robot can mimic a human-like response, such as displaying mannerisms indicating the feeling of happiness, without experiencing the corresponding emotion.

10. Conclusion

With the growing sophistication of robots, it is becoming increasingly likely that the technology will have a profound influence on human behavior. Equipped with the knowledge of affective computing and other disciplines, roboticists could exert profound power over the humans that interact with robots. Thus, we sought to explore a possible design pathway whereby robots would seek to nurture a user's empathy towards other human beings and more specifically, nudge the user towards the performance of charitable acts. Given that numerous computing devices are already covertly and overtly shaping user behavior, the use of robots to accomplish similar goals is a topic area that warrants significant ethical scrutiny, a discussion we aimed to start here.

References

ABC News (2007) Parents: Cyber bullying led to teen's suicide.
http://abcnews.go.com/GMA/story?id=3882520&page=1. Accessed 6 May 2016

Alba A (2015) Mattel's talking Hello Barbie doll raises concern over children's privacy. New York Daily News. http://www.nydailynews.com/news/national/mattel-barbie-raises-concern-children-privacy-article-1.2151019. Accessed 19 May 2016

Arkin R (2014a) Theory of mind models for robotic mediation of stigma in patient-caregiver relationships. 2014 Conference of the International Association for Computing and Philosophy (IACAP 2014). Thessaloniki, Greece

Arkin RC (2014b) Ameliorating patient-caregiver stigma in early-stage Parkinson's Disease using robot co-mediators. Proceedings of the AISB 50 Symposium on Machine Ethics in the Context of Medical and Health Care Agents. London, UK

Arkin R, Fujita M, Takagi T, Hasegawa R (2003) An ethological and emotional basis for human-robot interaction. Robotics and Autonomous Systems 42:191-201

Banks MR, Willoughby LM, Banks WA (2008) Animal-assisted therapy and loneliness in nursing homes: Use of robotic versus living dogs. Journal of the American Medical Directors Association 9(3):173-177

Bloom P (2015) The dark side of empathy. The Atlantic. http://www.theatlantic.com/science/archive/2015/09/the-violence-of-empathy/407155/. Accessed 20 June 2016

Bohannon J (2016) Government 'nudges' prove their worth. Science 352(6289):1042

Borenstein J, Arkin R (2016) Robotic nudges: The ethics of engineering a more socially just human being. Science and Engineering Ethics 22(1):31-46

Brooks A, Arkin RC (2007) Behavioral Overlays for Non-Verbal Communication Expression on a Humanoid Robot. Autonomous Robots 22(1):55-75

Carroll N (2011) On some affective relations between audiences and the characters in popular fictions. In: Coplan A, Goldie P (eds) Empathy: Philosophical and Psychological Perspectives. Oxford University Press, New York, pp 162-184

Chang W-L, Sabanovic S (2015) Studying assistive robots in their organizational context: Studies with PARO in a nursing home. HRI'15 Extended Abstracts Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts, 227-228

Coplan A, Goldie P (eds) (2011) Empathy: Philosophical and Psychological Perspectives. Oxford University Press, New York

Dolby N (2014) The future of empathy: Teaching the millennial generation. Journal of College and Character, 15(1):39-44

Dreyfus H (2004) Nihilism on the information highway: Anonymity versus commitment in the present age. In Feenberg A, Barney D (eds) Community in the Digital Age: Philosophy and Practice. Rowman & Littlefield, Maryland, pp 69-81

Endenburg N, van Lith HA (2011) The influence of animals on the development of children. The Veterinary Journal 190(2):208-214

Eyal N, Hoover R (2013) Hooked: How to Build Habit-Forming Products. Nir Eyal

Ferenbok J, Mann S, Michael K (2016) The changing ethics of mediated looking. IEEE Consumer Electronics Magazine 5(2):94-102

Fogg BJ (2003) Persuasive technology: Using computers to change what we think and do. San Francisco, Morgan Kaufman

Ham J, Midden C (2013) A persuasive robot to stimulate energy conservation: the influence of positive and negative social feedback and task similarity on energy consumption behavior. Int. J. Soc. Robot 6(2):163-171

Ham J, Spahn A (2015) Shall i show you some other shirts too? The psychology and ethics of persuasive robots. In: Trappl R (ed) A Construction Manual for Robots' Ethical Systems. Switzerland, Springer International Publishing, pp 63-81

Harris T (2016) How Technology Hijacks People's Minds—from a Magician and Google's Design Ethicist, http://www.tristanharris.com/2016/05/how-technology-hijacks-peoples-minds%e2%80%8a-%e2%80%8afrom-a-magician-and-googles-design-ethicist/. Accessed 12 June 2016

Hausman DM, Welch B (2010) Debate: to nudge or not to nudge. The Journal of Political Philosophy 18(1):123-136

Hern A (2016) Facebook, YouTube, Twitter and Microsoft sign EU hate speech code. The Guardian, https://www.theguardian.com/technology/2016/may/31/facebook-youtube-twitter-microsoft-eu-hate-speech-code. Accessed 31 May 2016

Hoffman G, Zuckerman O, Hirschberger G, Luria M, Shani Sherman T (2015). Design and evaluation of a peripheral robotic conversation companion. Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction, pp. 3-10

Hornyak T (2014) An AI milestone: Chatbot passes Turing Test by posing as 13-year-old boy. PCWorld, http://www.pcworld.com/article/2361220/computer-said-to-pass-turing-test-by-posing-as-a-teenager.html. Accessed 12 May 2016

Konrath S, O'Brien E, Hsing C (2011) Changes in dispositional empathy in American college students over time: A meta-analysis. Personality and Social Psychology Review 15(2):180-198

Kosner AW (2014) Hooked: How to make habit-forming products, and when to stop flapping. Forbes, http://www.forbes.com/sites/anthonykosner/2014/02/17/hooked-how-to-make-habit-forming-products-and-when-to-stop-flapping. Accessed 12 June 2016

Levy D (2007) Love and Sex with Robots. Harper Perennial

Lin P, Abney K, Bekey GA (eds) (2014) Robot Ethics: The Ethical and Social Implications of Robotics. The MIT Press

Masunaga S (2016) Here are some of the tweets that got Microsoft's AI Tay in trouble. Los Angeles Times, http://www.latimes.com/business/technology/la-fi-tn-microsoft-tay-tweets-20160325-htmlstory.html. Accessed 6 June 2016

Mettler K (2016) Gen Con, major gaming convention, has more female than male speakers for the first time ever, making some gamers grumpy. The Washington Post, https://www.washingtonpost.com/news/morning-mix/wp/2016/05/18/gen-con-major-gaming-convention-has-more-female-than-male-speakers-for-the-first-time-ever-and-some-gamers-arent-happy-about-it/. Accessed 18 May 2016

Moshkina L (2011) An Integrative Framework of Time-Varying Affective Robotic Behavior. Ph.D. Dissertation, School of Interactive Computing, Georgia Institute of Technology

Moshkina L, Arkin RC (2005) Human perspective on affective robotic behavior: A longitudinal study. Proc. IROS-2005. Calgary, Canada

Nye DE (2004) Technological prediction: A Promethean problem. In: Sturken M, Thomas D, Ball- Rokeach SJ (eds) Technological Visions: The Hopes and Fears that Shape New Technologies. Temple University Press, pp 159-176

Orwell G (1950) 1984. Penguin Publishing Group

Park S, Moshkina L, Arkin RC (2010) Mood as an affective component for robotic behavior with continuous adaptation via learning momentum. Proc. 10th IEEE-RAS international Conference on Humanoid Robots (Humanoids 2010). Nashville, TN

Pettinati M, Arkin RC (2015) Towards a robot computational model to preserve dignity in stigmatizing patient-caregiver relationships. 2015 International Conference on Social Robotics (ICSR 2015). Paris, France

Prinz JJ (2011) Is empathy necessary for morality? In: Coplan A, Goldie P (eds) Empathy: Philosophical and Psychological Perspectives. Oxford University Press, New York, pp 211-229

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

Scasselati B, Admoni H, Mataric MJ (2012) Robots for use in autism research. Annual Review of Biomedical Engineering 14:275-294

Shim J, Arkin RC (2015) An intervening ethical governor for a robot mediator in patient-caregiver relationships. International Conference on Robot Ethics (ICRE 2015). Lisbon, Portugal

Stephan A (2015) Empathy for artificial agents. International Journal of Social Robotics 7:111-116

Stueber K (2014) Empathy. In: Zalta EN (ed) Stanford Encyclopedia of Philosophy. http://plato.stanford.edu/archives/win2014/entries/empathy/. Accessed 1 June 2016

Swick K (2005) Preventing violence through empathy development in families. Early Childhood Education Journal 33(1):53-59.

Swick K (2001) Nurturing decency through caring and serving during the early childhood years. Early Childhood Education Journal 29(2):131-137

Thaler RH, Sunstein CR (2008) Nudge: Improving Decisions About Health, Wealth, and Happiness. Yale University Press, New Haven

Thomas F, Johnston O (1981) The Illusion of Life: Disney Animation. Hyperion

Turkle S (2011) Alone Together: Why We Expect More from Technology and Less from Each Other. Basic Books

Turkle S (1995) Life on the Screen: Identity in the Age of the Internet. Simon and Schuster Paperbacks, New York

Turkle S (2015) Reclaiming Conversation: The Power of Talk in a Digital Age. Penguin Press

Turkle S (2005) Relational artifacts/children/elders: The complexities of cybercompanions. Cognitive Science Society, Stresa, Italy

Vinik D (2015) Obama's effort to 'nudge' america. Politico, http://www.politico.com/agenda/story/2015/10/obamas-effort-to-nudge-america-000276. Accessed 1 July 2016

Wallach W, Allen C (2009) Moral machines: Teaching robots right from wrong. Oxford University Press, Inc, New York

Weigel M (2016) Flirting with humanity: The search for an artificial intelligence smart enough to love. New Republic, https://newrepublic.com/article/133034/flirting-humanity. Accessed 19 May 2016

Whitby B (2012) Do you want a robot lover? The ethics of caring technologies. In: Lin P, Abney K, Bekey GA (eds) Robot Ethics: The Ethical and Social Implications of Robotics. MIT Press, pp 233-248

The White House (2015) Executive Order -- Using Behavioral Science Insights to Better Serve the American People, https://www.whitehouse.gov/the-press-office/2015/09/15/executive-order-using-behavioral-science-insights-better-serve-american. Accessed 1 July 2016

Zinn W (1993) The empathic physician. Archives of Internal Medicine 153(3):306-312

Zuckerman O, Hoffman G (2015) Empathy objects: Robotic devices as conversation companions. Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction, pp 593-598