

Social Robots and Empathy: The Harmful Effects of Always Getting What We Want

Camylle Lanteigne

An important feature of the 21st century so far seems to be the increasing personalization and individualization of numerous facets of life as we know it. One area where this has arisen is with news outlets. On the one hand, the number of sources from which one can get their news, and the relative diversity of the positions one might encounter, are tremendously greater than a few decades ago—and this is good, to the extent that an aggregate of different views is often more accurate than a one-sided story. On the other hand, things are not so simple. While there are more voices, individuals still prefer the ones that fit their own pre-existing beliefs best.¹ This can sometimes lead to paradigm-shifting claims, such as the existence of “alternative facts,” which Kellyanne Conway famously coined in response to the discrepancy between the recorded number of people who attended Trump’s inauguration and Sean Spicer’s claims on the matter. One of the central issues here is that individuals with widely different viewpoints are now entitled to their own *facts*, making it superfluous to try and understand someone else’s take on any aspect of life because *opinions* are open to discussion, not facts. In other words, each person can go on believing that their understanding of the world is right, as they need not ever change their mind or revise their beliefs. This, I contend, can have dire consequences on empathy, as it eliminates any reason for trying to see why someone might hold a certain point of view, especially someone whom we consider “different” in terms of nationality, race, political affiliation, gender, or religious beliefs, for instance.

¹ Knobloch-Westerwick, Silvia, and Jingbo Meng. “Looking the Other Way: Selective Exposure to Attitude-Consistent and Counterattitudinal Political Information.” *Communication Research* 36, no. 3 (June 2009): 426–48. <https://doi.org/10.1177/0093650209333030>, p. 443.

This increased individualization has also deeply shaped technology, as we very rapidly went from owning a family computer—or even using computers at the library—to each owning a laptop, a tablet, a mobile phone. Many people are now concerned with the still misunderstood effects of technology on our brains, on our ways of life, and even on empathy. For one, Charles Harvey believes that our obsession with technology is not only making us intellectually shallow—as evidence shows we are having more and more trouble sustaining deep concentration—but also *emotionally* shallow. This is because, on the one hand, we have come to *learn and understand* in constantly more interrupted and more hurried ways—we read many short articles while listening to music instead of reading one book in silence—as well as *interact* with others in more interrupted and more hurried ways—we send our friends many scattered messages instead of having a long, uninterrupted conversation.² However, “the capacity to empathize with others, to recognize the felt otherness of the other, requires a *history of prolonged contact and concentration, a slow and unhurried processes.*”³ These two elements—our increased shallowness and the simultaneous requirement for depth in empathy—according to Harvey, will make us more apt for romantic relationships with robots than with humans.⁴ In light of the above, my main thesis for this essay will be that empathizing with social robots will only make us even less empathetic towards other humans since social robots remove the need to empathize with other human beings in the same way alternative facts do: by making it possible to have *exactly* what we want, without compromise.

In arguing for this thesis, I will first show that we are effectively able to empathize with social or sociable robots—understood as “a physically embodied, autonomous agent that

² Harvey, Charles. “Sex Robots and Solipsism: Towards a Culture of Empty Contact.” *Philosophy in the Contemporary World* 22, no. 2 (2015): 80–93. <https://doi.org/10.5840/pcw201522216>, p. 87.

³ Harvey, p. 87. Emphasis added.

⁴ Harvey, p. 89.

communicates and interacts with humans on a social level.”⁵ In other words, it is a robot we should be able to relate to, and empathize with.⁶ Secondly, I will present the hypothesis that the Uncanny Valley phenomenon, which can disrupt empathy towards robots, is tied to disgust, which underlies dehumanization. Thirdly, I will argue that social robots are much easier to empathize with, since they can be programmed to meet a person’s every desire. Following from this, I will also argue that empathizing with social robots will strengthen moral disgust towards humans. Fifthly, as in the uncanny valley hypothesis mentioned above, the moral disgust we will come to feel towards other human beings will lead to their dehumanization, a dehumanization caused—ironically—by the fact that humans will appear to be *too* human. Finally, I will end by putting forward two specific cases where this dehumanization of humans could play out in the near future.

To begin with, there is neuroscientific evidence that we are able to empathize with social robots. The fundamental reason why this is possible is because of the “mirror neurons” in our brains. Indeed, mirror neurons are believed to be responsible for the phenomenon of resonance, which Chaminade *et al.* define as “the mechanism by which the neural substrates involved in the internal representation of actions, as well as emotions and sensations, are also recruited when perceiving another individual experiencing the same action, emotion or sensation.”⁷ Simply put, if I see a person smiling, the neurons in my brain that would fire if I were smiling do so here, too. Chaminade *et al.*, in attempting to understand the effect of anthropomorphizing robots, found that the neurons in the parts of the brain associated with a specific emotion also fired when a human

⁵ Darling, Kate. “Extending Legal Protection to Social Robots: The Effects of Anthropomorphism, Empathy, and Violent Behavior towards Robotic Objects.” In *Robot Law*, by Ryan Calo, A. Froomkin, and Ian Kerr, 213–32. Edward Elgar Publishing, 2016. <https://doi.org/10.4337/9781783476732.00017>, p. 215.

⁶ Breazeal, Cynthia L. *Designing Sociable Robots*. MIT Press, 2004, p. 1.

⁷ Chaminade, Thierry, Massimiliano Zecca, Sarah-Jayne Blakemore, Atsuo Takanishi, Chris D. Frith, Silvestro Micera, Paolo Dario, Giacomo Rizzolatti, Vittorio Gallese, and Maria Alessandra Umiltà. “Brain Response to a Humanoid Robot in Areas Implicated in the Perception of Human Emotional Gestures.” *PLOS ONE* 5, no. 7 (21 July 2010): e11577. <https://doi.org/10.1371/journal.pone.0011577>, p. 2.

saw a *robot* face expressing emotions such as joy, anger and disgust.⁸ Therefore, if robots are effectively able to convey emotions to humans, as Itoh *et al.* have also found⁹, this opens the possibility for humans to be able to empathize with social robots, as empathy minimally requires that one be able to perceive the emotions of the other. Indeed, empathy is usually defined as the ability to put oneself in another's shoes, to understand one's situation as if we ourselves were in it. In the words of Belzung: "It is a response triggered by the emotional state of the other, but it is also the recognition and the more or less precise understanding of his (her) mental states."¹⁰ In the case of social robots, even if they do not really feel emotions, the emotional behaviour they display pushes us to imagine the mental states they *would* have, and empathize accordingly, as examples throughout this text will show.

Additionally, there is a growing body of evidence of humans anthropomorphizing social robots—that is, talking or acting as if the robot had emotional states or cognition while knowing very well it does not. Take, for instance, people who own a Pleo, an endearing robotic baby dinosaur, and blog about it on pleoworld.com. One person writes that their Pleo was "born"¹¹ when they first turned it on, points out that it has a "fun-loving and energetic"¹² personality, while someone else points out that their Pleo felt "tired" or "scared."¹³ What is more, there is significant

⁸ Chaminade *et al.*, p. 10. For further evidence of our mirror neurons firing when presented with actions from a robot, see Gazzola, V., G. Rizzolatti, B. Wicker, and C. Keysers. "The Anthropomorphic Brain: The Mirror Neuron System Responds to Human and Robotic Actions." *NeuroImage* 35, no. 4 (1 May 2007): 1674–84. <https://doi.org/10.1016/j.neuroimage.2007.02.003>.

⁹ Itoh, K., H. Miwa, M. Matsumoto, M. Zecca, H. Takanobu, S. Roccoella, M. C. Carrozza, P. Dario, and A. Takanishi. "Various Emotional Expressions with Emotion Expression Humanoid Robot WE-4RII." In IEEE Conference on Robotics and Automation, 2004. TExCRA Technical Exhibition Based., 35–36, 2004. <https://doi.org/10.1109/TEXCRA.2004.1424983>, p. 36.

¹⁰ Belzung, Catherine. "Empathy". *Journal for Perspectives of Economic, Political, and Social Integration*; Lublin 19, no. 1–2 (2014): 177–91. <http://dx.doi.org/10.2478/v10241-012-0016-4>, pp. 178-179.

¹¹ Jacobsson, Mattias. "Play, Belief and Stories about Robots: A Case Study of a Pleo Blogging Community". In RO-MAN 2009—The 18th IEEE International Symposium on Robot and Human Interactive Communication, 232–37. Toyama, Japan: IEEE, 2009. <https://doi.org/10.1109/ROMAN.2009.5326213>, p. 2.

¹² Jacobsson, p. 3.

¹³ Jacobsson, p. 4.

evidence not only for the anthropomorphizing of Pleos, but also for *empathy* being directed towards them. Still on the pleoworld.com blog, one person, when their Pleo's skin started to peel and the paint to come off, wrote: "Poor Pleo!!! It's like she's sick!!!! =(".¹⁴ Similarly, Kate Darling recounts an incident where, in the context of an experiment, small groups of people were given Pleos and then asked to "tie up, strike, and 'kill' [them]".¹⁵ When this happened, some people went as far as physically protecting the Pleo and interfering with other members of their group that were going to hurt the Pleo which, when "hurt," whimpers and displays pain behaviour.¹⁶ In a different experiment, research participants presented with a video of a Pleo being tortured reported they felt, among other things, "empathetic concern" for it after viewing the video.¹⁷ Clearly, individuals feel empathy for the Pleo, as not only do they assign it characteristics like pain and suffering, but they *themselves* suffer when confronted with the "suffering" of the Pleo. Thus, humans can empathize with social robots.

Let us now turn to the uncanny valley phenomenon. First coined by Japanese roboticist Masahiro Mori, the "uncanny valley" refers to the idea that

humanlike objects like certain kinds of robots elicit emotional responses similar to real humans proportionate to their degree of human likeness. Yet, if a certain degree of similarity is reached emotional responses become all of a sudden very repulsive. The corresponding recess in the supposed function is called the uncanny valley.¹⁸

Additionally, the uncanny valley deepens when the humanlike object can move.¹⁹ While the uncanny valley phenomenon is fairly well-known by now, there is still some disagreement as to

¹⁴ Jacobsson, p. 4.

¹⁵ Darling, p. 222.

¹⁶ Darling, p. 223.

¹⁷ Rosenthal-von der Pütten, Astrid M., Nicole C. Krämer, Laura Hoffmann, Sabrina Sobieraj, and Sabrina C. Eimler. "An Experimental Study on Emotional Reactions Towards a Robot". *International Journal of Social Robotics* 5, no. 1 (1 January 2013): 17–34. <https://doi.org/10.1007/s12369-012-0173-8>, p. 29.

¹⁸ Misselhorn, Catrin. "Empathy with Inanimate Objects and the Uncanny Valley". *Minds & Machines* 19, no. 3 (August 2009): 345–59. <https://doi.org/10.1007/s11023-009-9158-2>, p. 345.

¹⁹ Mori, M., K. F. MacDorman, and N. Kageki. "The Uncanny Valley [From the Field]". *IEEE Robotics Automation Magazine* 19, no. 2 (June 2012): 98–100. <https://doi.org/10.1109/MRA.2012.2192811>, p. 99.

what exactly is the nature of the repulsion we feel, or why it arises. One promising hypothesis on the matter, put forward by Angelucci, Graziani, and Grazia Rossi, begins by positing that the revulsion that characterizes the uncanny valley interferes with the process of “recognition,” the “willingness to interact with other beings in an *empathic* way.”²⁰ According to this same hypothesis, *disgust* is at the basis of our revulsion towards the humanlike object that falls into the uncanny valley, whether it is a zombie or a creepy-looking humanoid robot. While disgust can be understood as physical disgust, “intended as a purely physiological reaction to various contaminants,” it can also refer to moral disgust, “intended as a state of intellectual repugnance.”²¹ What is crucial about these two “kinds” of disgust is that they have, in fact, through evolution, become *intertwined*. Therefore, when physical disgust is triggered—which the authors believe happens in the uncanny valley phenomenon—our moral disgust is simultaneously elicited, meaning that empathy can very difficultly be directed towards something we feel physically disgusted by.²² They then define *dehumanization* as “the more or less conscious and more or less intentional denial of either humanness *tout court*, or single human traits to other members of our species.”²³ In conclusion, they suggest that dehumanization may be the bridge that links the feeling of disgust to the uncanny valley phenomenon.²⁴

While the Angelucci, Graziani, and Grazia Rossi do not go further into their exploration of this hypothesis, I want to highlight the crucial link they make between the arousal of disgust, the inhibiting of empathy that follows from disgust, and the posited dehumanization that seems to be the result. Indeed, when we experience disgust towards someone or something, we do not tend to

²⁰ Angelucci, A., Graziani, P., & Grazia Rossi, M., “The Uncanny Valley: A Working Hypothesis” in Nørskov, M. (Ed.). (2016). *Social robots: boundaries, potential, challenges*. Farnham, Surrey, UK ; Burlington, VT: Ashgate, p. 125. Emphasis added.

²¹ Angelucci *et al.*, p. 133.

²² Angelucci *et al.*, p. 133.

²³ Angelucci *et al.*, p. 133.

²⁴ Angelucci *et al.*, p. 134.

empathize with the disgusting entity, as these two emotions are in conflict.²⁵ For instance, if someone commits a gruesome murder, as I find this morally disgusting, I will not tend to empathize with this person or feel for them if they seem to be having a difficult time adapting to life in prison. I might nonetheless empathize with another felon in the same situation whom I do not judge to be morally repulsive, like someone with a single minor drug-dealing offence, for example. Now, while social robots may not any time soon lead us to be *physically* disgusted by human beings, I strongly believe that they may help generate *moral* disgust towards humans which will also lead to the inhibition of empathy and dehumanization that Angelucci, Graziani, and Grazia Rossi talk about, but towards humans this time.

But before, I will show why social robots are easier to empathize with. First off, one area where the possibility of social robots taking the place of humans was explored is with love robots. Indeed, one predicted consequence of the anthropomorphizing of and empathizing with social robots is that some humans may eventually *fall in love* with them.²⁶ This raises a myriad of ethical issues, but the one I wish to focus on concerns the claim that humans will in fact *prefer* relationships with robots to those with humans.²⁷ The reason why this is possible is because robots can essentially be programmed to fulfill every requirement and desire of their user—physically, intellectually and emotionally—making it much more attractive than a human being that has flaws or quirks we cannot simply programme away.²⁸ Now, it may seem quite extreme to think that we would ever become so shallow as to prefer a robot to a human being simply because the robot won't argue with us, be in a bad mood, or forget to take the trash out. Yet, as it was mentioned

²⁵ Duhaime-Ross, Arielle. "Empathy and Disgust Do Battle in the Brain". Scientific American. Accessed 18 July 2019. <https://www.scientificamerican.com/article/empathy-and-disgust/>.

²⁶ Levy, David N. L. *Love + Sex with Robots: The Evolution of Human-Robot Relations*. 1st ed. New York: HarperCollins, 2007 in Harvey, Charles. "Sex Robots and Solipsism: Towards a Culture of Empty Contact". *Philosophy in the Contemporary World* 22, no. 2 (2015): 80–93. <https://doi.org/10.5840/pcw201522216>, p. 81.

²⁷ Harvey, p. 80.

²⁸ Sullins, p. 400.

earlier, evidence suggests that this is, in fact, the direction we are heading in since we are *already* becoming less able to empathize with human beings due to the lack of deep and prolonged interaction that is necessary for empathy. Thus, if we cannot bring ourselves to empathize with our partner when they are curt with us because they have had a bad day, or when they are too busy to pick up the kids at school, then our ability to remain with this person and maintain a supportive and respectful relationship seems severely threatened. However, a social robot that is always in a good mood, always does what we ask it to, and happens to suit our fancy in all the other ways suddenly seems like a plausible romantic partner, as it does not present any disagreeable behaviour that would make it difficult for us to empathize with it. For these reasons, it seems that social robots are much easier to empathize with, and in times like these, when we are struggling to empathize, we may very well prefer the easier option.

So why exactly might our lack of empathy towards humans, conjointly with our newfound preference for the company of social robots, lead us to feel moral disgust towards other humans? First, remember that moral disgust is tied to intellectual repugnance. If one is repudiated by a person morally, they may very well think that the morals that guide this person's actions are repugnant in comparison with their own morals, which they believe are superior.²⁹ Now, a lack of empathy towards others, especially others we consider to be different from us for various reasons, is detrimental to how openly and tolerantly we respond to this person, and how we treat them.³⁰ Additionally, while social robots can be programmed to never disagree with us or upset us, they can also be designed to be no different from us and share the same ethnicity, religious beliefs, and

²⁹ Simply put, it does not make sense for one to hold moral values and live their life according to these while not thinking (at least implicitly) that these are the best.

³⁰ Butrus, Ninawa, and Rivka T. Witenberg. "Some Personality Predictors of Tolerance to Human Diversity: The Roles of Openness, Agreeableness, and Empathy". *Australian Psychologist* 48, no. 4 (2013): 290–98. <https://doi.org/10.1111/j.1742-9544.2012.00081.x>, p. 297.

moral preferences as us. Consequently, our already deficient empathy towards humans, mixed with the high desirability of social robots that are made to satisfy all our preferences, make it that our ability and incentive to overcome the moral disgust we may feel towards others become weaker and weaker. Thus, this is how moral disgust is promoted by our empathizing with social robots.

Subsequently, in relation to Angelucci's, Graziani's, and Grazia Rossi's suggestion that dehumanization may be what ties disgust to the uncanny valley phenomenon, it seems quite plausible that dehumanization follows from disgust. However, unlike the dehumanization that may occur in the uncanny valley phenomenon, the dehumanization of humans is here centred upon the inconsiderate and unempathetic treatment of human beings precisely because they are *too* human, too complex and unpredictable in comparison to the tailored-to-our-every-desire social robot. This overwhelming complexity, Harvey tells us, is due to the lack of empathy we have towards human beings, as this lack makes it more difficult for us to "tolerate the complexity of others."³¹ Therefore, the empathizing with social robots can only reinforce this impression of overcomplexity towards humans as the robots are anything but complex to empathize with since they can fit our every desire. Humans are hence treated in a dehumanizing way, all because they have become too complex, too human in the face of perfect robot companions.

Let us now consider two concrete instances—in addition to that of love robots—where, in the near future, social robots and our deficiency in empathy towards humans could contribute to our dehumanization as I have described it. First, as social robots are becoming increasingly common as caregivers or companions to children and the elderly alike, individuals and institutions that use them will want the robots to fit their own religious beliefs or "moral universe."³² This

³¹ Harvey, p. 89.

³² McBride, James. "Robotic Bodies and the Kairos of Humanoid Theologies". *Sophia*, 5 December 2017. <https://doi.org/10.1007/s11841-017-0628-3>, p. 4.

implies that what the robot will teach, how it will respond to humans, and even its physical appearance will have to meet the religious preferences of those who use them.³³ While religion is seldom talked about in the context of robots, I believe that it is here of fundamental importance since religious beliefs are, historically, extremely powerful and at the centre of many conflicts. In light of this, it seems especially important to promote tolerance and empathy between faiths. However, as I have argued above, our empathizing with social robots is leading us towards the opposite of this. Religious robots, I believe, will therefore contribute to diminishing tolerance, empathy, and even openness towards people of faiths different than our own by promoting companionship with a being that never risks unsettling our own beliefs or teaching us the value in others' beliefs. Consequently, because caregiver and companion social robots seem to be the first making their way into our lives, and because religious beliefs are so important to the people who hold them, "religious" social robots are likely to, in the near future, directly affect our ability to empathize with other human beings, especially those who have different religious beliefs than ours.

A second case that may, in the near future, affect our ability to empathize with humans concerns the physical abuse of social robots. The fear, here, is a bit different from what we have seen so far: in this case, it is *not* empathizing with the social robots around us that risks negatively affecting how we treat humans. Indeed, if one acts violently towards a social robot that doesn't feel pain but expresses pain behaviour, one may become desensitized to pain behaviour, and more easily act violently towards sentient creatures, like human beings, who *do* feel pain while displaying pain behaviour.³⁴ This would further inhibit our ability to empathize with human beings

³³ What is more, even if the robot is not explicitly religious, people will still want it to fit, for example, their core values, or even just their cultural background.

³⁴ Darling, p. 224.

as it disregards completely the interior life of a person and their ability to suffer, and I can hardly see how one can empathize if they do not recognize that the other has the same kinds of feelings as them. In light of this, it seems that both empathizing and not empathizing with social robots could have a negative effect on empathy towards humans, as empathizing strengthens our tendency to prefer what is tailored to our desires, while not doing so can make us less attuned to the interior life of others. In consequence, because the physical abuse of social robots can make us blind to the suffering of sentient creatures, and because desensitization to pain behaviour could arise from social robots that are already on the market (such as Pleos), the physical abuse of social robots is likely to, in the near future, affect humans' ability to empathize with other humans and encourage their dehumanization.

To conclude, I have argued that, because we are already losing our ability to empathize with humans due to their complexity, we will increasingly turn towards social robots that are tailored to our every desire for companionship. This will only further accentuate the decline of our empathy for other human beings, and allow for moral disgust to set in, especially towards individuals we consider to be different from us. Because disgust underlies dehumanization, moral disgust towards our fellow humans will lead us to dehumanize them—that is, to forego treating them in the respectful and empathetic way we would expect one to treat a fellow human being—as this is something we will reserve for social robots who do not disagree with us or act in ways that don't suit us. Thus, the fatal flaw of humans will have been that they are *too human*, *too other* for the increasingly self-centred and digital world we live in now.

It is true that these harmful effects I argue social robots will bring are not unprecedented—people have preferred and continue to prefer to associate with those that are most like them, or those they best get along with. Nonetheless, I strongly believe there is something groundbreaking

and deeply unnerving about being able to programme your caregiver, your friend, or your lover to be exactly the way you want them to be; something that has so far been impossible in a purely human world. In consequence, if we want to preserve the social cohesion that is already fading, we must give serious thought as to how we will continue to integrate social robots into our society and simultaneously ensure humans aren't excluded from it.