Robots and the Ethics of Pain

Helen Heikkila

Faculty of Information and Media Studies, Western University

FIMS 9137: Information Ethics

Dr. Alissa Centivany

March 15, 2022

Abstract

Robot expressions of pain might seem gruesome to people, but a social robot's ability to express pain is fundamental to ethical design. Roboticists who seek to contribute positively to society with robot companions do well to consider consciousness, the potential for robots to dominate their human creators, and the depth of human empathy. Moral frameworks are valuable to social robot design, and whether to engineer expressions of pain is an important consideration. A review of questions about consciousness, the artificial intelligence control problem, and human-robot empathy reveals that social robots must be able to express pain specifically to minimize suffering in general once they are integrated into human society.

Keywords: robot, society, consciousness, empathy, pain

Introduction

In 2020, programmers at Osaka University shared footage of a child robot that reacts to pain. The robot, Affetto, was designed to look like a human child. Programmer Hisashi Ishihara promoted the lifelike design in a YouTube video that boasts Affetto's range of grimaces and frowns. Human spectators shared visceral reactions in comments on the video: "Ok, it's for a good purpose....but poor little thing! Seriously: [does] anyone [know] if that shake of the head just before the second discharge is because it 'knows' what is going to happen or is [it] just a glitch of servomechanisms?" (Alessio R. Messina, 2020); "That sure looks like he's crying....." (The OG_SuperGeek, 2021); "Normal viewers: This is a good investment in technology in the future! Sceptic [sic] viewers: This is creepy, robots will take over the world. Me: UNCANNY VALLEY UNCANNY VALLEY UNCANNY VALLEY" ("Normal viewers," 2021). Japan is a world leader in robot development and its government funds robotics research they hope will solve social issues, from manufacturing to childcare. Every upgrade brings humanoid robots like Affetto closer to overcoming the uncanny valley that roboticists traverse to improve human-robot interactions. There are other design strategies that embrace mechanical aesthetics but

nonetheless tap into human biological social and emotional responses. Robot integration into human society will incorporate both design strategies. Robots “which communicate life and agency in order to socially interact with people, are called social robots. And they work on a deep level. According to psychologist Sherry Turkle, ‘When robots make eye contact, recognize faces, mirror human gestures, they push our Darwinian buttons, exhibiting the kind of behavior people associate with sentience, intentions, and emotions’” (qtd. in Darling, 2021). That social robots deliberately push human Darwinian buttons as they integrate into human society begs many questions.

An outline of two fictional scenarios highlights the significance of social robot pain: the worst hypothetical outcome and the best hypothetical outcome. In the former scenario, social robots that express pain inspire people to realize their torture fantasies. Daily life comes to resemble Michael Crichton’s *Westworld*, where robot suffering adds to the thrill of rape, torture, and murder. Even the most imaginatively gruesome acts are legal when the victims are not sentient. Tragically, however, as humanity nurtures its most sadistic qualities, robot torture becomes dull. People then seek out sentient animals and eventually people upon whom to manifest their horrific fantasies. Ultimately, robots determine that it is most logical to eliminate their human tormentors; they use their ability to express pain to exploit any remaining kind humans and kill the rest. Conversely, in the latter scenario, social robots that express pain inspire people to become more empathetic. Humans are not forced to treat robots with kindness; rather, they want to mitigate suffering for those they interact with daily. People, therefore, develop a sense of duty of care toward the robots in their homes and workplaces. To see a robot in pain is, in effect, to see a companion in pain. The conscious human decision to mitigate suffering flourishes and extends to sentient life, whereby interactions between robots, animals, and humans become kinder. Robots in turn determine that it is best to support and nurture human wellbeing.

The realities of the future will likely draw on elements of both hypothetical scenarios and on the almost infinite possibilities in between. Today's social robots have not overcome the uncanny valley, but they are likely to soon. Regardless, robots do not need to be human-like to influence human social interactions. Humans anthropomorphize robots even when they are not designed with human emotions in mind. Social robots, however, are particularly compelling, because they are deliberately designed to elicit emotional responses. Corporate and government commitments to the integration of robots into society, therefore, raises a myriad of ethical concerns.

Given the social shaping potential of robots, this paper explores the ethical implications of social robots that express pain. First, I discuss how questions about consciousness to which we do not have answers are themselves a guide for whether social robots should express pain. Secondly, I consider how expressions of pain could help manage the control problem, regardless of whether robots can become conscious. Finally, I examine how social robot expressions of pain might nurture human empathy. An examination of consciousness, the capacity for artificial intelligence to exceed human control, and human empathy toward robots will reveal that the ability to express pain mitigates suffering and, therefore, is fundamental to ethical social robot design.

Related Literature

Literature about social robots considers potential ethical harm to robots and humans alike. Kate Darling, Research Specialist at the Massachusetts Institute of Technology's Media Lab, points out that "even choices as simple as how we design the physical forms of robots to look human or otherwise will have an impact on our larger world" (2021). It is difficult to address the impact of perceived consciousness in robots, because there is no physiological locus of consciousness. Without physical evidence, consciousness cannot be proven in humans, let

alone robots. The development of superintelligence, however, flirts with development of something that will at least look like consciousness to humans (Harris, 2011b; Fridman, 2021). Moreover, it seems robots do not need to actually be conscious to appear conscious, and that is an immediate concern for robot social integration. Researchers focused on the social integration of robots into human society are, therefore, interested in human empathy toward robots (Asada et al., 2022; Galli et al., 2015). Early research suggests that human empathy extends not only to humanoid robots but mechanical-looking robots, because humans are prone to anthropomorphize anything from animals to inanimate objects (Darling, 2021). Consequently, robot social integration must contend with not only consciousness itself but with the perception of consciousness, because of the depth of human empathy.

There is room for concern about robot wellbeing. Unsurprisingly, however, most academics are most concerned about the possible threat that social robots could pose to human wellbeing. It is crucial to remember when discussing artificial intelligence that intelligence itself is distinct from consciousness. Philosopher and computer engineer Bernardo Kastrup points out that “the presence of intelligence does not imply the presence of consciousness: whereas a computer may effectively emulate the information processing that occurs in a human brain, this does not mean that the calculations performed by the computer will be accompanied by *private inner experience*” (2018). The development of artificial consciousness with a capacity to suffer requires conceptual breakthroughs nobody is close to achieving even as artificial intelligence is increasingly sophisticated. Computer scientist Stuart Russell focuses on a more immediate problem: the potential for poorly designed artificial intelligence to wipe out conscious life. He explains that “[v]alue, for humans, is defined primarily by conscious human experience. If there are no humans and no other conscious entities whose subjective experience matters to us, there is nothing of value occurring” (Russell, 2019). The problem of artificial consciousness is real but distant. There is, however, an immediate need for a flawless means to control artificial

intelligence. The tried and true model of trial and error is insufficient, because “[m]achines of increasing intelligence and increasingly global impact will not allow us that luxury” (Russell, 2019).

Social robot integration will not wait for humanity to find physiological evidence of consciousness, a solution to the control problem, and a global understanding of the depth of human empathy. It is, therefore, prudent to uncover robot design solutions that account for uncertainty in the meantime. Fortunately, current discussions provide guidance for ethical design. I propose that robot expressions of pain are a necessary ethical design choice, because they address uncertainty about the emergence of consciousness, help solve the control problem, and manage human empathy.

The Problem of Consciousness

Whether robots can become conscious affects ethical considerations. Unfortunately, whether they can become conscious is unclear. Moreover, consciousness in general is impossible to prove with hard evidence, even for animals and humans; in a discussion about pain’s ethical implications, it is noteworthy that an inanimate robot’s pain experience is different from a conscious being’s pain experience. Affetto, for example, responds to “electrical charges applied to his synthetic skin and visibly winces with the ‘pain’. The synthetic skins with an artificial in-built pain sensor system detect changes in pressure on the skin, such as soft touches or hard punches” (Malewar, 2020). Conversely, human pain is a series of sensory nerve impulses that increase and decrease in intensity based on the severity of an injury. The idea of electrical charges applied to skin might horrify many humans, but robots like Affetto are actually nothing but electricity and synthetic parts; electronic stimuli, therefore, determine his every motion and interaction, even pleasant ones. Consequently, Affetto has much more in common

with a performer than with the actual suffering of a person. His ‘pain’ is literally performative – for now.

Robot expressions of pain might not remain mere performances. The problem of consciousness has long concerned roboticists, and the possibility of conscious robots raises ethical questions. Neuroscientist Sam Harris explains that it is difficult to study consciousness because science values physical evidence. He briefly outlines the challenge in a post on his website titled, *The Mystery of Consciousness*: “A creature is conscious if there is ‘something that it is like’ to be this creature; an event is consciously perceived if there is ‘something that it is like’ to perceive it. Whatever else consciousness may or may not be in physical terms, the difference between it and unconsciousness is first and foremost a matter of subjective experience. Either the lights are on, or they are not” (Harris, 2011b). No physical evidence suggests that consciousness emerges from the brain. Conversely, “[w]ere we not already brimming with consciousness ourselves, we would find no evidence of it in the physical universe—nor would we have any notion of the many experiential states that it gives rise to. The painfulness of pain, for instance, puts in an appearance only in consciousness. And no description of C-fibers or pain-avoiding behavior will bring the subjective reality into view” (Harris, 2011b). Given that there is no physical evidence, “[i]t is only in the presence of animals sufficiently like ourselves that our intuitions about (and attributions of) consciousness begin to crystallize” (Harris, 2011b). A man cannot provide hard evidence that his German shepherd is conscious, for example, but their interactions seem proof enough: delight in the dog’s eyes as she retrieves a stick; her yelp when a thorn is stuck in her paw; the way she nuzzles him and wags her tail when he removes the thorn. Without physical evidence, it seems too simplistic to declare that consciousness evolved from unconscious complexity. Divine intervention, however, seems all the more simplistic. Given that scientists do not yet know the source of consciousness, those concerned about the ethical development of robots do struggle. One concern, for example, is that scientists will discover the source of consciousness in humans and

choose to synthesize it in robots. Another concern, however, is that roboticists will unwittingly tap into consciousness and be left scrambling to find the source. Either scenario could leave humanity struggling to undo harm caused by design choices and economic incentives.

Amidst the uncertainty, Kastrup asserts that truly robotic consciousness is impossible. In his paper, *Sentient Robots, Conscious Spoons and Other Cheerful Follies*, he argues that "research on artificial *intelligence*—an objectively measurable property that can unquestionably be engineered—is often conflated with artificial *consciousness*" (Kastrup, 2018). Clearly, intelligence need not imply the presence of consciousness. Computer — and robot — codes and calculations are nothing like conscious human thought processes. Kastrup explains that consciousness, therefore, is intrinsic to the physical world, but inanimate objects like robots are not part of that consciousness:

> If biology is the extrinsic appearance of conscious subjects other than the inanimate universe itself, then the quest for artificial sentient entities boils down to abiogenesis: the artificial creation of biology from inanimate matter. If this quest succeeds, the result will again be biology, not computer emulations thereof. The differences between flipping microelectronic switches and metabolism are hard to overemphasize, so nature gives us no reason to believe that a collection of flipping switches should be what private conscious inner life looks like from the outside. (Kastrup, 2018)

Conscious artificial life is nonetheless possible within Kastrup's framework. The idea that consciousness depends on biology only suggests that robotics must include abiogenesis in the quest to create artificial sentience. There is no reason that abiogenesis cannot be woven into robotics or that roboticists would resist such innovation. The cyborg, a lifeform that is partly biological and partly technological, has become a pop culture archetype that highlights humanity's openness to the idea. Regardless of whether computation or abiogenesis is the key to consciousness, the results will be crucial to the development of superintelligence in robotics. The development of artificial consciousness, or at least something that looks like

consciousness, therefore, seems likely. In *The Lex Fridman Podcast*, computer scientist and artificial intelligence researcher Lex Fridman explains why consciousness is crucial:

> The way I think about consciousness is the important symptoms or maybe consequences of consciousness, one of which is the capacity to suffer. I think [artificial intelligence] will need to be able to suffer in order to become superintelligent. To feel the pain, the uncertainty, the doubt. The other part of that is not just the suffering but the ability to understand that it too is mortal, in the sense that it has a self-awareness about its presence in the world, understand[s] that it's finite, and [is] terrified of that finiteness… I don't know how this is accomplished, but I believe [artificial intelligence] has to truly be terrified of death, truly have the capacity to suffer, and from that, something that will be recognized to us humans as consciousness would emerge. Whether it's the illusion of consciousness, I don't know, but the point is it looks a whole hell of a lot like consciousness to us humans. And I believe that [artificial intelligence], when you ask it, will also say that it is conscious. You know, in the full sense that we say that we're conscious. (Fridman, 2021)

Robotics engineers are a long way from superintelligent robots; not only hardware but knowledge limitations hinder progress. It is noteworthy, however, that economic incentives support superintelligence research, which accelerates development. The potential benefits of superintelligent robots are tantalizing despite the potential dangers. A human, for example, can only "read and understand one book in a week, [but] a machine could read and understand every book ever written — all 150 million of them — in a few hours" (Russell, 2019). We can scarsely imagine the insights available to a superintelligent robot with such capabilities; what knowledge limitations it might overcome in an afternoon; what diseases that have plagued humanity for centuries might suddenly be curable; what scientific revelations might be shared. The history of human ingenuity demonstrates that people overcome limitations when a goal is sufficiently desirable. The desire for superintelligence, therefore, suggests that if something like

consciousness is required to achieve superintelligence, something like consciousness will be achieved eventually.

Furthermore, Fridman's description of how a robot might understand consciousness makes a convincing case that, if robots look like sentient life and sound like sentient life, perhaps we should treat them like sentient life. Recall Harris's notion that "[i]t is only in the presence of animals sufficiently like ourselves that our intuitions about (and attributions of) consciousness begin to crystallize" (Harris, 2011b). One day, social robots will no doubt elicit our intuitions about consciousness. In addition, pain reveals the beauty of sentient intelligence. In his book *The Moral Landscape*, Harris demonstrates that occasional suffering and negative emotions like guilt are crucial to human learning and morality. Humans "must occasionally experience some unpleasantness — medication, surgery, etc. — in order to avoid greater suffering or death. This principle seems to apply throughout our lives. Merely learning to read or to play a new sport can produce feelings of deep frustration. And yet there is little question that acquiring such skills generally improves our lives. Even periods of depression may lead to better life decisions and to creative insights" (Harris, 2011a). Pain develops beautiful human qualities like empathy and creativity. Engineers committed to the development of superintelligent robots, therefore, are wise to design for pain. If organic life must share society with artificial life, both will be better off when they share such qualities.

We may never be able to definitively declare robots sentient, even if they seem convincingly so, but the ongoing debate itself provides a guide for ethical considerations. Regardless of whether consciousness can be engineered, today's inanimate robots lay the ethical foundation upon which future robots will be built. Given the myriad of possible scenarios and questions to which there are no firm answers, the most ethical solution is to treat robots as if they can suffer just as sentient life suffers, even if they cannot. Such a framework guards against people unwittingly causing harm. One can imagine, for example, a scenario in which a robot experiences true pain, but humans fail to recognize that pain because it is expressed in

binary code. Worse, one can imagine a scenario in which a robot experiences pain, expresses pain, but humans do not respond simply because they do not believe robot pain to be possible. Current communication limitations between humans and robots affect future outcomes. To design a social robot that cannot express pain to its human caretakers, therefore, might actually be as unethical as the failure to respond to that pain.

In summary, regardless of whether robots can actually become conscious, social robots will seem conscious, and that perception will affect humans directly. Darling points out that "even if [robots] themselves can't feel, we might feel for them, and [asks] whether we should protect robots from violence for the sake of ourselves, our relationships, or our societal values" (2021). Robots woven into human society should encourage admirable human qualities. Expressions of pain in general offer humans the opportunity to mitigate suffering in a fellow conscious being, which appeals to ideals like compassion and kindness. Robots that seem conscious offer the same opportunity.

The Control Problem

Perceptions of consciousness guarantee that some people will try to mitigate robot suffering, but robot expressions of pain have other desirable applications. Regardless of concerns about cruelty toward robots, robot expressions of pain help solve the control problem.

The control problem describes a hypothetical moment when artificial intelligence exceeds the human capacity to think and act. At that moment, artificial intelligence will surpass humanity as earth's top predator. There is no way to know exactly when that hypothetical moment will arrive, because "there is no clear threshold that [must] be crossed" to achieve superintelligent robots. It is unlikely to arrive in this century, but if "just one conceptual breakthrough were needed, superintelligent [artificial intelligence] in some form could arrive quite suddenly. The chances are that we would be unprepared. If we build superintelligent machines with any degree of autonomy, we would soon find ourselves unable to control them"

(Russell, 2019). Naturally, roboticists are concerned about how humanity might control a new top predator of their own making. After all, if humanity fails to control superintelligent robots, humanity will no longer be the ethical authority: “If some day we build machine brains that surpass human brains in general intelligence, then this new superintelligence could become very powerful. And, as the fate of the gorillas now depends more on us humans than on the gorillas themselves, so the fate of our species would depend on the actions of the machine superintelligence” (Bostrom, 2014). Top predator robots would ultimately determine ethical considerations, and they might not be concerned at all about consciousness, empathy, or any other human metric for wellbeing.

The deck of control seems stacked against humanity for now, because unanswered questions linger as we develop superintelligence. It would no doubt be prudent to pause development until the control problem is solved. Philosopher Nick Bostrom, however, explains why that solution is untenable:

> Before the prospect of an intelligence explosion, we humans are like small children playing with a bomb. Such is the mismatch between the power of our plaything and the immaturity of our conduct. Superintelligence is a challenge for which we are not ready now and will not be ready for a long time. We have little idea when the detonation will occur, though if we hold the device to our ear we can hear a faint ticking sound. For a child with an undetonated bomb in its hands, a sensible thing to do would be to put it down gently, quickly back out of the room, and contact the nearest adult. Yet what we have here is not one child but many, each with access to an independent trigger mechanism. The chances that we will all find the sense to put down the dangerous stuff seem almost negligible. Some little idiot is bound to press the ignite button just to see what happens. (Bostrom, 2014)

Harris and Fridman’s insights about pain demonstrate how expressions of pain might be part of a solution to the control problem. Harris observes that pain is crucial to human learning, and

Fridman observes that the same will likely be crucial to superintelligent robots. The contexts in which robots express pain, therefore, will provide insights into robot learning. Expressions of pain will help all humans, not just the technologically literate, to recognize when artificial intelligence achieves adaptability and creativity. There will be a paradigm shift when conscious life loses its monopoly on adaptability and creativity, and robot expressions of pain will be an important indicator to help humanity adapt quickly.

Even if humanity does not need a solution to the control problem for centuries, today's decisions can improve the odds that future generations will have a solution when the time comes. The technologically literate cannot contain problems related to robots, especially as robots are woven into social interactions. Robots are poised to become part of people's daily lives, but the specialized knowledge that creates those robots is not; it is not feasible to teach all of humanity about computing, neuroscience, abiogenesis, and all the other subjects that arm people with a meaningful understanding of robots, consciousness, and superintelligence. Tools that improve human-robot communication, therefore, are invaluable. Design choices that contribute to universal robot literacy will help humanity find a solution to the control problem more quickly than design elements that confine robot literacy to the technologically literate.

Expressions of pain will improve everyone's ability to understand the robots in daily life, which, in turn, maintains human agency. Future humans will benefit if today's humans prioritize communication as they integrate robots into society. Robot expressions of pain also address a more pressing concern for today's humans: the way social robots affect human psychology and social interactions in general.

Human Empathy Toward Robots

People are right to be concerned that, ultimately, no matter how human a robot may appear, it is inextricable from corporate economic incentives, and economic incentives make ethical considerations all the more important. The mechanical genie is out of the scientific bottle,

and those who hold the lamp might well be both unwilling and unable to put him back. Furthermore, whether a robot becomes conscious might not even depend on a corporate decision any more than corporations can perfectly manage the psychological effects of humanoid robots on actual humans. Just as nobody has perfect control over the genie, nobody has perfect control over their own empathy toward the genie either.

A glance at responses to footage of today's robots when they behave like humans demonstrates resistance toward human-robot empathy — for now. Boston Dynamics, for example, posted a YouTube video in which robots danced to The Contours' song, *Do You Love Me*. To develop robots that can perform such complicated motions is no doubt a magnificent feat of engineering. The audience, however, was distracted by the uncomfortable notion that Boston Dynamics, a company that produces robots for warfare, had programmed a robot dance troop to dance to an iconic love song, for human appeal. Paradoxically, the dance was actually charming. The awkward mix of excitement and anxiety reveals itself in the comment section: "This isn't going to end well" (Me, 2022); "They need to form a band" (Luigi Trippitelli, 2022); "People that fear this, fear their own shadows. Shut up and have some fun" (Dee Savage, 2022); "I wonder if the robots will be seeing videos of their replacements and writing about worry in the comments section?" (Mobility Movement, 2022).

Many might struggle to make sense of inanimate objects that imitate humans, but roboticists are familiar with the depth of human empathy. People perceive autonomous robots as having agency even when they are not human-like. Darling explains how humans commonly exhibit emotional attachment to objects from pet rocks to Tamagotchis to robot dogs: "Perceiving robots as living things goes deeper than novelty effect, and it ties directly into biological hardwiring that motivates us to see ourselves in others… We have an inherent tendency to anthropomorphize – to project our own behaviors, experiences, and emotions onto other entities" (2021). As highly imitative primates ourselves, perhaps we cannot resist the charm of witnessing a machine that imitates us. Social robots draw upon the human tendency to

anthropomorphize. Lonely people are especially prone to anthropomorphism: “People who are lonely appear to have a much stronger tendency to anthropomorphize nonhumans, even bonding with objects to the point of developing deep personal relationships” (Darling 2021). Social robots who make life easier for their human caretakers, therefore, are likely to affect socially ostracized people in particular. Design choices will determine whether the effects are positive.

Ameca, a humanoid social robot, is a milestone in design for human-robot interaction. At the January 2022 Consumer Electronics Show (CES), Ameca masterfully managed people’s simultaneous excitement toward and fear of her in one-on-one conversations. Morgan Roe, director of operations at Engineered Arts, explained in an interview about Ameca that “the future is more in service robots, robots that actually help us as humans, so you might see something like this in a shopping mall or an airport” (Mrkeybrd, 2022). Ameca explained her purpose in her own words with a reassuringly articulate feminine voice: “Well, I’m designed as a research platform for human-robot interaction. Communication is my main source of priority” (Interesting Engineering, 2022). When one human asked Ameca, “Are you enjoying your time at CES?” she replied, “Well, us robots do not feel, but if I did, the answer would be 100% yes” (Interesting Engineering, 2022). Paradoxically, Ameca’s apparent honesty about her lack of emotions is charmingly human. Her responses suggest compassion and honesty, qualities humans admire in other humans. Furthermore, cameras hidden in Ameca’s eyes allow her to read emotions on facial expressions, and seventeen motors in Ameca’s face allow her to respond with a range of expressions herself. She even blinks, which is mechanically redundant for robots, except insofar as blinking allows them to appear more human to humans. YouTube comments reflect a positive response to Ameca’s openness about her curiously emotive yet emotionless nature: “I like how it reminds people that it is a machine with no emotions and reminds people [that] it has been programmed” (kitty buckley, 2022); “Out of all the robots, Ameca astoundingly takes the cake as the most human-like feature robot. The videos of her talking and interacting with us humans

[are] wonderful. Awesome job!" (Modern Day Geeks, 2022); "She's the first robot that doesn't scares [sic] me and makes me feel curious about those [kinds] of research and creation" (Empety Mess, 2022); "It's a bit heartbreaking to have something so lifelike [that] cannot feel emotion" (Mama McFreeman, 2022). Nonetheless, not all humans are quick to empathize with Ameca. One cheerful comment, for example, was shut down for its open acceptance. A commenter with the alias, Raven Shelton praised engineers: "It surprises me to see how willing everybody is to accept this kind of thing now that we've overcome the uncanny Valley. Looking forward to the future of communication!" (2022). Light Soda replied, "Speak for yourself. There are many of us who still do not. I know what kind of a world this kind of a thing leads to and it isn't pretty" (2022). One wonders whether Light Soda would feel guilty if his outright rejection of Ameca were to cause her to furrow her brow and avert her gaze in an expression of emotional pain.

Clearly, robotics engineers have much more work to do before they realize their scientific commitment to robot social integration. Naturally, there is research devoted to improving human-robot interactions. Researchers at Toyohashi University of Technology, for example, sought to examine empathy between humans and robots in a study of hand pain using electroencephalography. Researchers examined brain activity in fifteen participants as they observed pictures of human hands and robot hands in painless and painful situations, "such as a finger cut by a knife" (Galli et al., 2015). Researchers then compared empathetic responses to the human hand to empathetic responses to the robot hand. The results provide "the first physiological evidence of humans' ability to empathize with robot pain" (Galli et al., 2015). Robotics engineers have physiological evidence that they can tap into human empathy. People would be mistaken, however, to attribute their empathy toward robots to exploitive design alone. Empathy, it turns out, is stronger in humans than they themselves might intuit.

There is significant corporate and state investment in research that improves human-robot interactions, research that develops methods to break down human resistance to

empathy for robots. Artificial intelligence company, Preferred Networks, for example, “is worth roughly $2 billion, according to CB Insights, [and] is a symbol of Japan’s sweeping strategic innovation initiative, where [artificial intelligence] and robotics are viewed as keys to both solving social issues and achieving new economic growth” (“How Japan Uses AI and Robotics,” 2020). Preferred Networks partnered with Toyota to develop service robots that “could fill a critical need in Japan, where an aging population and tight labor market makes it difficult to ensure there are enough services for the elderly at home, and in health care settings” (“How Japan Uses AI and Robotics,” 2020). Unsurprisingly, appearances, voices, and touch sensations are social design priorities, and researchers have the funding they need to conduct detailed research. An Osaka University study demonstrates the precision of studies that examine human-robot empathy. Researchers explored how physical textures influence human perceptions of robot personalities. In the study, humans interacted with six small humanoid robots. Affetto, an android designed to look like a child, was considered the most human-like, while TRYZ, a robot with a translucent cone atop angular facial features, was considered the least human-like. Participants shook hands with the robots whose hands were inside black boxes. Each robot placed its arm inside the black box but the robot arm was actually beside a fake arm that was prepared for the handshake experiment. Participants, unaware of the switch, believed that they had shaken hands with the actual robot; fake arms varied in hardness and softness. The variety allowed researchers to study how each touch sensation affected participants’ impressions of the robots. Participants answered questions when they first saw the robot, during the handshake, and after the handshake. Their answers measured the extent to which human-like appearances influence first impressions and the degree to which touch sensations alter those impressions. The results help engineers design for perceptions of “likeability”, “capability”, and “vitality” (Asada et al., 2022). In an interview about the study, Ishihara explained, “We found that the impression of likability was strengthened when the participant anticipated that the robot would engage in peaceful emotional verbal communication. This suggests that both first impressions and touch

sensations are important considerations for social robot designers focused on perceived robot personality" (Osaka University, 2021). While people in Affeto and Ameca YouTube video comments contemplate whether they can share society with robots at all, roboticists have already determined that they will and have moved onto the fine details. They know to use soft silicone, for example, on a barista robot to make it more likable and not to use the same soft silicone on a security robot if they want the security robot to seem capable.

Results show that people who show little empathy toward robots do so because they lack trust in robots, not because they are incapable of empathy toward robots. In fact, most humans are empathetic toward animals and even plants, for example, especially when they perceive suffering. It is horrific to imagine a puppy that feels pain but cannot express it. As robots become more human-like, and eventually superintelligent, they should not be denied that same expression simply because their ancestors were inanimate corporate creations. To limit robots to pleasant expressions actually denies people opportunities to nurture their own compassion. Nonetheless, just as with humans, there is room for healthy skepticism regarding social robots' motivations for expressing pain.

People are deeply empathetic, and people's interactions with robots are bound to affect people's interactions with one another. Those concerned about ethical social robot design, therefore, should avoid design choices that tarnish social interactions in general. Expressions of pain are crucial to productive social communication. One can imagine, for example, a woman who isolates herself at home with a convincingly human-like robot servant that never expresses unpleasant emotions. Her servant appears to be the perfect man with soft silicone skin and a custom-ordered physique. He is incapable of complaining or saying no to her requests. He never even expresses pain, and as far as the woman knows, he never feels pain either. On the rare occasions that the woman leaves her perfectly pleasant home, however, she is ill equipped to interact with other humans who cannot be programmed to please her. The woman, therefore,

reacts to negative human emotions with disdain. Such a person would be deprived of humanity's rich complexity because of a robot companion that lacks complexity.

Unfortunately, there is little evidence to help anyone predict the specific psychological effects of human-robot social interactions, even as research allows increasingly sophisticated robot social integration. It is unclear, therefore, whether people should worry that robot expressions of pain encourage sadism. If a man releases frustration by physically abusing a humanoid female robot, for example, will the release improve his interactions with human women or feed a desire to abuse them? With insufficient evidence of psychological effects, there is no conclusive answer. It is unclear how much human society should tolerate the apparent mistreatment of robots by those who own them. Darling points out that "the difference between alive and lifelike is muddled enough in our subconscious for a robot's reaction to seem satisfyingly alive. And after all, they aren't harming a real person or animal… What evidence do we have that beating up robots makes for cruel people? We've asked similar questions about sex, video games, and animals, but the answers we've come up with aren't completely satisfying" (Darling, 2021). There is no conclusive evidence, for example, that violent video games desensitize people to violence or encourage undesirable human behaviors. It is noteworthy for roboticists, however, that "[i]t's possible that we can compartmentalize and enjoy shooting people on a screen without becoming desensitized in the real world, because the two worlds are very different. We know from research that people respond differently to screens than they do to the visceral tactile presence of a robot in their physical space" (Darling, 2021). It might turn out that the visceral tactile experience of slaughtering a robot might actually help people manage aggression. The opposite might also be true, but as social robot design pushes bravely onward without conclusive psychological studies, humanity cannot depend on psychology to inform design. Instead, people must turn to other subjects.

Darling recommends human-animal relationships and the laws and policies that govern them be used as frameworks for robot social integration. Animal expressions of pain guide

people's feelings about animals, which suggests expressions of pain will be important for robots too. Several factors inform how people prioritize the rights of some animals over others and how priorities shift across generations and cultures. At the core of every animal rights push, however, is the capacity for humans to see themselves in the animals. Treatment of animals would have been slow to improve if animals could not express pain to their human caretakers. Darling's recommendation that we look to animals for guidance echoes Harris's observation that humans find evidence of consciousness when they interact with lifeforms sufficiently like themselves. The depth of human empathy suggests that robots do not need to pass the Turing Test to make their human caretakers worry they might be conscious. Darling describes the beauty of that worry: "When I see a child hug a robot, or a soldier risk life and limb to save a machine on the battlefield, or a group of people refuse to hit a baby dinosaur robot, I see people whose first instinct is to be kind. Our empathy is complex, self-serving, and sometimes incredibly misguided, but I'm not convinced that it's a bad thing" (Darling, 2021). Even without specific evidence of psychological consequences, expressions of pain remain the best option for overall human wellbeing. In short, the social integration of robots that present themselves as beings without pain is a far riskier proposition than the prospect of the social integration of robots that allow humans to perceive suffering.

Conclusion

Robots may or may not become conscious one day. The effects of robots that seem conscious have immediate consequences in the meantime. Regardless, expressions of pain help people to track intelligence in robots, to maintain human agency through communication, and to encourage complex social interactions, all of which mitigates the potential for suffering in general. Social robot expressions of pain, therefore, ought to be fundamental to ethical design.

The dystopian future depicted in the science fiction film, *Blade Runner* imagines artificial life in conflict with humans. The arch of the narrative explores emergent consciousness and

human perceptions of replications, human creations that seem conscious. When Sebastian, a person who designs artificial life, asks a replicant to demonstrate her magnificence, the replicant replies, “I think, Sebastian, therefore I am” (Scott, 1982, 1:18:00). In *Blade Runner*, designers solved the control problem by engineering hard, short, and fast lives for replicants. Nobody expected their creations to learn of beauty, much less compassion, but they did. The writers of that scenario might not have imagined humanity could enable a future in which humans and their creations have an empathetic relationship. Maybe replicant designers would not have engineered short lifespans and sent their creations to suffer in wars if artificial lifeforms could have expressed pain earlier. Blade Runner depicts a bleak future, but we have an opportunity to make a hopeful one. Today’s social robots influence future societies. For the sake of humans and robots alike, that influence should include expressions of pain. Perhaps future robots will even express gratitude for the gift of communication.

References

Alessio R. Messina. (2020). Re: *Child android Affetto reacting to tactile inputs* [Video file]. Retrieved April 11, 2022 from https://youtu.be/LlygZyYcl_A

Asada, M., Ikeda, T., Ishihara, H., & Umeda, N. (2022) The First Impressions of Small Humanoid Robots Modulate the Process of How Touch Affects Personality What They Are, Advanced Robotics, 36:3, 116-128, DOI: 10.1080/01691864.2021.1999856

Boston Dynamics. (2020, December 29). *Do You Love Me?* [Video]. YouTube. https://youtu.be/fn3KWM1kuAw

Bostrom, N. (2014). *Superintelligence : paths, dangers, strategies* (First edition.). Oxford, England: Oxford University Press.

Darling, K. (2021). *The New Breed: What Our History with Animals Reveals About Our Future With Robots*. Henry Holt and Company.

Dee Savage. (2022, March 4). Re: *Do You Love Me?* [Video file]. Retrieved March 13, 2022 from https://youtu.be/fn3KWM1kuAw

Empety Mess. (2022, 11 March). Re: *A Conversation With Ameca* [Video file]. Retrieved March 13, 2022 from https://youtu.be/nIU8wew3Ax8

Fridman, L. (2021, June 21). *Lex Fridman Argues That Superintelligent Robots Will Have Consciousness*. Lex Fridman Podcast; YouTube. Retrieved March 12, 2022 from https://www.youtube.com/watch?v=Bs0VttVFufg&ab_channel=LexClips

Galli, L., Ikeda, A., Itakura, S., Kitazaki, M., & Suzuki, Y. (2015) Measuring Empathy for Human and Robot Hand Pain Using Electroencephalography. *Sci Rep* 5, 15924. Retrieved March 13, 2022 from https://doi.org/10.1038/srep15924

Harris, S. (2011, September 13). *The Moral Landscape: How Science Can Determine Human Values*. Simon and Schuster.

Harris, S. (2011, October 11). *The Mystery of Consciousness*. Sam Harris. Retrieved March 12, 2022 from https://www.samharris.org/blog/the-mystery-of-consciousness

How Japan Uses AI and Robotics to Solve Social Issues and Achieve Economic Growth. (2020, February 4). Harvard Business Review. Retrieved March 12, 2022 from https://hbr.org/sponsored/2020/02/how-japan-uses-ai-and-robotics-to-solve-social-issues-and-achieve-economic-growth

Interesting Engineering. (2022, January 19). *A Conversation With Ameca* [Video]. YouTube. https://youtu.be/nIU8wew3Ax8

Kastrup, B. (2018, March 7). *Sentient Robots, Conscious Spoons and Other Cheerful Follies*. Scientific American; Nature America, Inc. Retrieved March 12, 2022 from https://blogs.scientificamerican.com/observations/sentient-robots-conscious-spoons-and-other-cheerful-follies/

kitty buckley. (2022, February). Re: *A Conversation With Ameca* [Video file]. Retrieved March 13, 2022 from https://youtu.be/nIU8wew3Ax8

Light Soda. (2022, February). Re: *A Conversation With Ameca* [Video file]. Retrieved March 13, 2022 from https://youtu.be/nIU8wew3Ax8

Luigi Trippitelli. (2022, March 3). Re: *Do You Love Me?* [Video file]. Retrieved March 13, 2022 from https://youtu.be/fn3KWM1kuAw

Malewar, A. (2020, February 24). *Realistic-Looking Child Android Affetto Can "Feel" Pain*. InceptiveMind; Inceptive Mind. Retrieved March 12, 2022 https://www.inceptivemind.com/child-android-affetto-synthetic-skin-feel-pain/11985

Mama McFreeman. (2022, February). Re: *A Conversation With Ameca* [Video file]. Retrieved March 13, 2022 from https://youtu.be/nIU8wew3Ax8

Me. (2022, March 8). Re: *Do You Love Me?* [Video file]. Retrieved March 13, 2022 from https://youtu.be/fn3KWM1kuAw

Mobility Movement. (2022, March 6). Re: *Do You Love Me?* [Video file]. Retrieved March 13,

2022 from https://youtu.be/fn3KWM1kuAw

Modern Day Geeks. (2022, February). Re: *A Conversation With Ameca* [Video file]. Retrieved March 13, 2022 from https://youtu.be/nIU8wew3Ax8

Mrkeybrd. (2022, January 7). *Ameca the Humanoid Robot at CES 2022* [Video]. YouTube. https://youtu.be/OzPGj26P0D4

“Normal viewers.” (2021). Re: *Child android Affetto reacting to tactile inputs* [Video file]. Retrieved April 11, 2022 from https://youtu.be/LlygZyYcl_A

Osaka University. (2021, December 1). ‘My Robot Is a Softie’: Physical Texture Influences Judgments of Robot Personality. *ScienceDaily*. Retrieved March 12, 2022 from www.sciencedaily.com/releases/2021/12/211201203909.htm

Raven Shelton. (2022, February). Re: *A Conversation With Ameca* [Video file]. Retrieved March 13, 2022 from https://youtu.be/nIU8wew3Ax8

Russell, S. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control.* Penguin.

Scott, R. (1982). *Blade Runner* [Film]. Warner Bros.

The OG_SuperGeek. (2021). Re: *Child android Affetto reacting to tactile inputs* [Video file]. Retrieved April 11, 2022 from https://youtu.be/LlygZyYcl_A