Oxford Studies in Experimental Philosophy, Volume 2

Tania Lombrozo (ed.) et al.

https://doi.org/10.1093/oso/9780198815259.001.0001

Published: 2018 **Online ISBN:** 9780191853012 **Print ISBN:** 9780198815259

CHAPTER

11 On the Matter of Robot Minds

Brian P. McLaughlin, David Rose

https://doi.org/10.1093/oso/9780198815259.003.0012 Pages 270–312

Published: March 2018

Abstract

One reason it matters whether phenomenally conscious robots will soon be forthcoming is that such robots would have moral rights. The view that they are on the horizon often rests on a certain philosophical view about consciousness, one called "nomological behaviorism" in this chapter. The view entails that, as a matter of nomological necessity, if a robot had exactly the same patterns of dispositions to peripheral behavior as a phenomenally conscious being, then the robot would be phenomenally conscious. The chapter experimentally investigates whether the folk think that certain (hypothetical) robots made of silicon and steel would have the same conscious states as certain familiar biological beings with the same patterns of dispositions to peripheral behavior as the robots. The findings provide evidence that the folk largely reject the view that silicon-based robots would have the sensations that they, the folk, attribute to the biological beings in question.

Keywords: Artificial Intelligence, sentience, phenomenal consciousness, robo-ethics, human-level intelligence, dispositions to peripheral behavior, analytical behaviorism, nomological necessity, nomological behaviorism

Subject: Moral Philosophy

Collection: Oxford Scholarship Online

1. Robot Minds Would Matter

In 1999, in Seattle, Washington, the American Society for the Prevention of Cruelty to Robots (ASPCR) was formed. A central tenet of the ASPCR is that sentient robots would have "unalienable rights."[1] The ASPCR acknowledges that there are no sentient robots on the planet, but claims that given the pace of technological advances, they are "much closer than previously thought." The mission of the ASPCR is "to ensure the rights of all artificially created sentient beings." The Society's objective is "to outline a robotic bill of rights, and eventually establish a lobbying board to further these rights on the body politic."

A sentient robot would indeed have inalienable rights. It would have moral rights. Sentience confers a kind of moral status. Sentient beings are moral patients. Their sentience is a moral consideration in any interaction with them.[2] If, for instance, a robot could feel pain or suffer in some way, it would deserve humane treatment. A legal system should protect its moral rights.

p. 271 It is also true that if a robot were to have the kind of mental abilities that would make it a moral agent, then it would have moral duties and obligations, and could be held morally accountable.[3] What those mental abilities are, exactly, is not a settled matter. Neither sentience nor genuine intelligence is enough. Many animal species are both sentient and intelligent, yet, of known species, only our own includes moral agents. We think the requisite mental abilities for moral agency are higher-order intentional states, together with the kind of mental sophistication that we normally find only in normal human beings past a certain age. If a robot had such mental abilities, it would be a moral agent, and so subject to moral censure, merit, and praise.[4]

What matters to whether a being is a moral agent or patient is just what, if any, mental abilities the being has. Moral status is conferred by mental abilities alone. That is one reason it matters deeply whether a robot has mental abilities.

There is a basic general principle here that can be stated as a supervenience thesis: There can be no difference with respect to general status as a moral agent or moral patient without a difference in mental abilities.[5] Robots that are both moral agents and moral patients would be full participants in the sphere of moral relations. At present, all known persons are human beings. But although such robots would not be human beings, they would be persons.

The ASPCR likens itself to the American Society for the Prevention of Cruelty to Animals (ASPCA). In 1823, the philosophy of mind and ethics began to join hands in support of the moral rights of animals when
p. 272 Jeremy Bentham made his now famous remark: "The question is not, Can ↳ they reason? Nor, Can they think?, but, Can they suffer?" (1823: ch. 27, fn.). The idea that sentient beings are moral patients eventually started to spread, and in 1866, in New York City, the ASPCA was formed.

When the ASPCA was formed, there were, as there are today, many suffering and mistreated animals. At present, there are no artifacts that suffer. There are no artificial moral patients or agents, because there are no sentient or genuinely intelligent artifacts. That is a point with which the ASPCR agrees. But if, as the Society anticipates, robots with sentience are on the way in the not too distant future, then the ASPCR's objective is urgent—how urgent depends on how soon such robots are to be expected if research continues unabated. Has the time come to join the cause of the ASPCR?

2. Mental Words Matter

If a robot were sentient, it would have moral rights. It is also true that if, as a panexperientialist might claim, a piece of jello is sentient, then it too would have moral rights. Indeed, if a witch cast a spell that made a broomstick sentient, then the broomstick would have moral rights. However, jello doesn't have mental abilities. And no one can cast a spell that makes a broomstick sentient. Whether a robot could be sentient is quite another matter. The ASPCR takes the pace of Artificial Intelligence (AI) research to be such that we should expect sentient robots sooner rather than later. If the Society is right about that, its mission is truly urgent indeed.

It was reported in *The New York Times* (November 17, 2003) that Hans Moravec, a leading roboticist at Carnegie Mellon University, stated: "I'm confident we can build robots with behavior that is just as rich as human being behavior." He added: "You could quiz it as much as you like about its internal mental life, and

it would answer as any human being." Moravec wasn't just making an off-hand comment to the press. In his *Encyclopedia Britannica Online* article, "Robots" (2003), he stated that if the current pace of robotic development continues, "robots are likely to parallel the evolution of vertebrate intelligence to the human
p. 273 level, and probably beyond, within fifty years."[6] In his *Scientific American* article (Moravec, ↳ 2009), he predicted that robot intelligence will surpass our own before 2050. If Moravec's prediction is correct, there will be robots with human-level intelligence or greater within 33 years.

If that prediction can be trusted, the urgency of the ASPCR's objective cannot be exaggerated, especially given the glacial pace at which countries deal with looming crises (think of the response to global warming). We should begin to act now. The robotics industry should be put on notice that if it ever develops robots with intelligence even approaching human-level intelligence, it will not own them, and will never be permitted to sell them. The sale of beings with intelligence approaching human-level intelligence would be slaving.[7] If the government failed to enact laws to prohibit the sale of such robots, even revolution would be called for were that the only way to stop it. To avoid this, we maintain that the government should shut down any research project with even an outside chance of developing such robots.

But to echo Ebenezer's question to the most fearsome of the three Ghosts: Are these shadows of things that will be or are they shadows of things that may be only?

Before sounding the alarm, it should be noted that in the 2003 version of his *Encyclopedia Britannica* article, "Robots," Moravec also says, in the context of reporting on the history of robotics: "By the late Twentieth Century automata controlled by computers could also think and remember."[8] If a being can actually think, then it is genuinely intelligent. Having the ability to think suffices for being genuinely intelligent. Thus, if we've had, for decades, machines that can think, the Holy Grail of the field of AI has in fact been in our hands for decades.

There are, to be sure, some truly impressive robots. The LEGO robot, made of Legos and a small computer, can solve the Rubik's cube from any starting arrangement to the completion of the puzzle in approximately 3.46 seconds. Mitsubishi-Heavy Industries Corporation's robot Wakmaru is capable of responding
p. 274 appropriately to over 10,000 spoken words, and can reliably detect up to 10 faces. *The Los Angeles Times* ↳ employs an algorithm, Quakebot, for producing news stories about earthquakes. Although it would be costly to do, we already have the existing technologies to put a mechanism that can execute Quakebot inside a humanoid robot, one made to look just like, say, Seymour Hersh, and to connect it with the robot's motor system. We could then have what we might call "the Hershbot" sitting in a cubicle, typing the story on a laptop.

AI has much to boast about. The Holy Grail of AI is not one of them. There are, currently, no thinking machines. Moreover, from what we have been able to gather from experts, we don't now see either sentient or genuinely intelligent robots on the horizon. And we regard it as an open question whether either sentient robots or robots with human-level intelligence will ever be technologically possible. Indeed, we regard it as an open question whether such robots are even nomologically possible (compatible with Mother Nature's laws). For all anyone now knows, such a robot might be like a machine that can transmit information faster than the speed of light, something that Mother Nature doesn't permit.

Mental words matter. If "thinking" or "sentience" is used in some proprietary AI sense not to be confused with the senses of those terms in ordinary English, then that should be made very explicit. Otherwise, it will inspire well-meaning groups such as the ASPCR.

If such AI mental talk is intended as literal, then those so talking should keep in mind that such prognostications are a double-edged sword. They might entice financial support, but the edge that cuts against AI research can do so in two ways. Promises can create expectations that won't be met. And there are promises no one should be allowed to keep.

In our discussion of robots in this chapter, we'll use both mental terms and moral terms literally. The issue of robot mentality is inseparable from the issue of robot-rights.

We stand in staunch opposition to the manufacture and sale of robots that have intelligence that even approaches human-level intelligence. We do so on the grounds that it would be slaving. We have, however, no objection at all to the sale of Mitsubishi-Heavy Industries Corporation's Wakmaru robots. They are now priced at $143,000. We only wish we could afford one ourselves.

We see no urgency in carrying out the ASPCR's mission. We don't think the United Nations should be
p. 275 mobilized. We feel no need to contact our congressperson or to alert the press. We don't even feel ↳ that we should join the ASPCR. The reason is that we don't share the ASPCR's assessment of the future of technology. Nanotechnology could perhaps develop to the point where a humanoid robot with the information storage capacity and information processing powers of IBM's Watson, and made to look just like, say, Lieutenant Commander Data of *Star Trek*,[9] could stand behind a podium with former *Jeopardy!*[10] Champions and win the grand championship. But despite its outward appearance, such a humanoid robot would no more have moral status than does Watson. It would be neither a moral patient nor a moral agent.

The world has already been put on alert about the (alleged) possibility of superintelligence, and the singularity. Nick Bostrom's (2014) book was a *New York Times* best seller. The singularity doesn't keep us awake at night. We believe that it now falls under the category of science fiction. Of course, what is at one time science fiction can later become science fact. Please, then, wake us up if, or when, there is a line of research that there is serious reason to think has even a remotely outside chance of developing superintelligence. We'll then do all that we can to work to help to stop it. But as for the current AI research of which we are aware, we're all for it, and enthusiastically encourage investment in it.

We've announced that the robots soon coming down the pipes will lack mental abilities. No doubt some members of the ASPCR will not feel reassured. No doubt some people will think that we are misinformed, naïve, or simply ignorant. Should cold water be thrown in our faces to wake us up from a dogmatic slumber? Are we naïve or ignorant in thinking that there will be no sentient or human-level intelligent robots in the foreseeable future? Unfortunately, we won't be able to settle those questions here. But later we'll clarify our position, and offer some considerations in defense of it.

For now, however, we want to turn to some genuinely pressing moral and social issues concerning extant robots and ones that will soon arrive, issues that arise even if, as we believe, such robots will be devoid of mentality.

3. Why It Really Matters That the Robots Are Coming

p. 276

There is a tsunami of robots on the way. It is estimated that by 2020, there will be 31.2 million robots on the planet. That could be a low-ball estimate. The South Korean government is aiming to have a robot in every Korean home by 2020 (Chamberlain, 2010). The population of South Korea reached 50 million in 2012. If the South Korean government comes anywhere in the ballpark of its goal, the number of robots worldwide will far exceed 31.2 million in 2020. In any case, to put the more conservative number of 31.2 million into perspective, in 2015 the population of Australia was 23.8 million and the population of Canada was 35.8 million. It could well turn out that in just three years from now, the number of robots on the planet will exceed, perhaps far exceed, the population of Canada.

Currently, the majority of the robots are Roombas, home vacuum cleaners developed by iRobot Corporation. But soon cloud-connected humanoid robots designed for service, including home care, will be entering the lives of people in affluent countries in unprecedented numbers. They'll provide valuable services indeed.

And they should be of no concern to the ASPCR. They will no more have moral rights than does a Roomba, which itself no more has moral rights than does an old-fashioned Sears and Roebuck push vacuum cleaner. They will lack mental abilities. Let's hope that such robots will be reasonably priced and widely available.

But even though such robots will be neither moral agents nor moral patients, they'll present a host of moral and social issues. The use of robots in war, and how a self-driving car should be programmed to respond to various kinds of dangerous situations, raise moral issues (Wallach and Allen, 2009). Then, there are such pressing social issues as the potential effects on employment. Smith and Anderson (2014) tell us that experts are divided over whether the flood of robots soon to come our way will increase or decrease employment in industrialized nations by the year 2025. But there are very serious grounds for worry. Industrial robots will lead to further decline of factory positions. There are now reasonably priced robots that can weld. We anticipate robots that can dig coal 24 hours a day, 365 days a week. It is estimated that
p. 277 there may be as many as 10 million self-driving cars on the road by as early as 2020 ↳ (Greenburg, 2016). Never mind potential loss of taxi driver jobs; in the US, more males are employed as truck drivers than in any other occupation (Smith and Anderson, 2014). Moreover, robots will soon fill many service-industry and white-collar jobs. Those are fields in which most women in the US are employed (Smith and Anderson, 2014).

The soon forthcoming humanoid service robots, specifically, present a plethora of moral issues that are all their own. The issues have to do not with our effects on them, since they'll lack mentality, and so can be neither harmed nor wronged. The issues have to do with their effects on us.

Eyes (real or artificial), faces, human-like or animal-like form, human-sounding voices, and matters such as biological motion influence our initial gut reactions to robots. We react viscerally to humanoid robots, which is different from the way we react to an iPad, even when the humanoid robot no more has genuine intelligence than does the iPad. It is a different experience to have a cloud-connected humanoid robot that seems to look into your eyes when answering your questions than it is to have Siri on your iPad answering your questions. Because of such visceral reactions, although it would be wrong to kick a Roomba in anger in front of a child, it would be far worse to kick or slap a humanoid robot in front of a child. That's not because either extant or soon forthcoming humanoid robots deserve humane treatment. If you owned one, there would be nothing morally wrong *per se* with you shutting it off, or even disassembling it and selling its parts. Slapping such a humanoid robot in front of a child would be wrong because of how observing such an interaction might affect the child. Also, we should now be concerned with such matters as how a child's growing up with a humanoid robot that seems to unquestionably obey its orders might influence the child's social interactions with other people the child regards neither as peers nor as having higher social standing. We don't want children to grow up thinking that they can just trump over such people.

Faces, eyes, and the like can make it appealing to interact with a robot. The roboticist Mashahiro Mori (1970/2012) anticipated, however, a phenomenon he dubbed "the uncanny valley." As robot appearance becomes more and more human-like, our gut reactions to robots change. We start to cease finding them appealing and start to enter the uncanny valley in which the robots seem creepy. Studies indicate that we
p. 278 initially react at a gut level as we would to a distorted human face or body ↳ (MacDorman and Ishiguro, 2006; Seyama and Nagayma, 2007; Tinwell, Grimshaw, Abdel Nebi, and Williams, 2011; Saygin, 2012). Especially disconcerting is the mismatch between their outward appearance and their behavior. It seems that because of their outward appearance we, at some level, expect them to behave like normal humans, expectations that are then dashed (Saygin, 2012). This failure of such unconscious expectations to be met is manifested in our consciousness by the cognitive feeling of creepiness. Realistic-looking humanoid robots, initially at least, can seem creepy.

Uncanny valley reactions have even been found in monkeys (MacPherson, 2009). The experiments with monkeys were done with animations. Psychologists find that we too have uncanny valley reactions to

animations (Burleigh, Schoenherr, and Lacroix, 2013). Many children who viewed the move *The Polar Express*[11] found the realistic appearance of the animations creepy. Fiona, in the movie *Shrek*,[12] was initially animated in a very realistic-appearing way, but children at the prescreening found the character so creepy that the production company had Fiona redone to make her appear less realistic.[13]

Although there are some very realistic-looking humanoid robots, the humanoid service and homecare robots soon on the way won't look much like humans. For sales reasons, industry doesn't want to manufacture creepy robots. To avoid inducing the feeling of creepiness and instead to induce warm and fuzzy feelings in us, robotic companies tend to produce cute, toy-like humanoid robots, such as, for instance, NAO, a cloud-connected humanoid robot developed by Aldebaran Robotics. NAO was inducted into the Carnegie Mellon University Robot Hall of Fame in 2012. A 2004 inductee, Honda's ASIMO (Advanced Step in Innovative Mobility), has been described as looking like a boy in a spacesuit, because of its size and shape. ASIMO robots could be made to look exactly like boys in regular boys' clothing. But, despite their truly impressive 17 degrees of freedom of movement, their movements are so different from normal human
p. 279 movements that if they ↳ looked like real boys in ordinary clothing, they would creep us out. Roboticists don't want to creep us out. They want to sell us robots. (At the time of this writing, a NAO is priced at $8,000, and an ASIMO, which costs a million dollars to make, can be leased for a month for $150,000. Again, we wish we had the money!)

Roboticists can now make extremely realistic-looking robots, but typically don't to avoid the uncanny valley. There is ongoing work to develop robots that rise above the uncanny valley in that they move smoothly in ways that appear very human-like, and make sounds indistinguishable from a human voice. Work under the directorship of Hiroshi Ishiguro at the Intelligent Robotic Laboratory at Osaka University in Japan deserves note, because of how realistic-looking and sounding the robots are. Ishiguro's robots, however, are mainly controlled by "teleporting"; that is, by receiving signals from a person at a terminal, rather than behaving autonomously, and so to that extent they are just fancy marionettes. Their voices, too, are the voices of the teleporter. Still, it looks as if research could very well result in humanoid robots that don't just look, move, and sound like humans, but that also extensively act autonomously.

It was 20 years from the time that a prototype ASMIO took its first step until ASIMOs were able to walk up and down stairs, run, hop on one foot, kick balls, shake hands, open jars, and pour coffee. If and when ASIMOs have smooth, fluid, human-like movements and human-sounding voices, and so rise above the uncanny valley, they may be made to look just like real boys, dressed in boys' clothing. Given that the robots will not be sentient and will have nothing even remotely approaching human-level intelligence (neither of which would be required for them to have very realistic human appearances, smooth human-like movements, and human-sounding voices), their sale wouldn't be slaving. There is no need to have Honda, the manufacturer of ASIMO, cease and desist. On the contrary, Honda should be encouraged. But it is important to recognize that such robots would raise concerns about humanoid–human interactions to an entirely new level.

Think in this connection of whether special orders will be taken to custom-build robots that look and sound exactly like a certain person (say, an ex-spouse, a co-worker, an acquaintance, or a celebrity). That, we believe, should be legally prevented—at the very least without the person in question's legal consent.

p. 280 A recent survey conducted in the UK by Martin Smith, the Head of the Mobile Robot Research Unit at the University of Middlesex, found that one in five people, a mixture of males and females, said they would be willing to have sex with a humanoid robot (reported in Lytton, 2014). Ian Yeoman and Michelle Mars (2012) predict that by 2050 Amsterdam's Red Light District will be staffed by android prostitutes, which, they point out, would remove health concerns, and the exploitation of human prostitutes. Consent is not an issue where so-called "sexbots" are concerned, because they are devoid of any mentality. But might a person's

interactions with sexbots affect how that person interacts with other people? If so, that is an enormously serious and deeply troubling concern.

Julie Carpenter (forthcoming) points out that sexbots could be used as sex surrogates by couples that have long-distance relationships, where one partner partly controls via teleporting the sexbot that is interacting with the other partner. Such robots would just be fancy sex toys, and so may seem harmless. But there are serious potential complications, especially as the sexbots rise above the uncanny valley, which they very well may do. To note one, David Levy (2008) predicts that by the year 2050, not only will many people be having sex with humanoid robots; many people will fall in love with them and want to marry them. The state of Massachusetts, he predicts, will be the first to legalize human–robot marriage.

Here we see shadows not of things that may be only, but of things that perhaps will indeed be (except, we think, for the part about Massachusetts). In fact, there are already actual cases of deep emotional attachments formed in robot–human interactions. There are actual cases of soldiers risking their lives on the battlefield to save bomb-detecting robots with which they have become emotionally attached (Carpenter, 2016).

Anthropomorphizing is thus already a serious concern.[14] It comes easily and naturally to us, and has been found in every culture (Eplsey, Waytz, and Caciopppo, 2007). In Europe in the Middle Ages, animal trials were common, and in some areas occurred as late as the nineteenth century. Pigs were often put on trial for murdering children (Evans, 1906: ch. 1). (Historians suspect that in many such cases the parents in fact killed the child, so as to have one fewer mouth to feed.) If the pigs were ↳ convicted, they were often put to death by hanging; but if the pig's act was deemed especially evil, it was buried alive instead (Evans, 1906: 138). Insects too were sometimes put on trial. There were cases in which captured locusts were put on trial for destroying crops (Evans, 1906: 135). As late as 1866, in Požega-Slavonia, after the conviction of some locusts that were captured red-handed eating crops, the locusts were "put to death by being thrown into the water with anathemas on the whole species" (Evans, 1906: 136). Sometimes, though, mercy was shown. In 1457, a sow and her piglets were tried for murdering a child (Evans, 1906: 154). Although the sow was convicted, the piglets were acquitted, on the grounds that they were not mature enough to make informed choices.

p. 281

Although anthropomorphizing is a serious concern, we don't now anticipate a repeat of scenarios like the ones above with humanoid robots (that are in fact neither sentient nor genuinely intelligent). Folk views have evolved too much since such dark times (right?). Still, there are many currently pressing moral and social concerns about humanoid–human interactions. We've listed some, but the list goes on, and on, and on. And no doubt issues will arise that we don't now anticipate. The point to note, though, is that the moral concerns are about how humanoid–human interactions will affect the humans involved. The robots could be damaged or destroyed, but neither harmed nor wronged.

Robot-ethicists are now engaged in examination of the issues, as well as in the study of what new laws it would be best to enact, given the kinds of robots already here, and that there is good reason to think will soon be coming (see the essays in Lin et al., 2011). There are pressing moral and social issues aplenty. Given the tsunami of robots on the way, a course devoted to the field of robot ethics (often called "robo-ethics") urgently deserves a place in the curriculum.

Robot-rights seem not to be much of an issue in the field, because it seems to be fairly widely assumed that the robots in question will be neither sentient nor genuinely intelligent. We think that's a very sensible assumption, at least for the foreseeable future. Members of the ASPCR and some members of the AI community, however, see matters differently. If they are right, then we think robot-rights should be at the very top of the agenda in robo-ethics. Indeed, the other issues, as important as they are, would pale in

comparison. Since the issue of robot-rights is inseparable from the issue of robot minds, let's return to the matter of robot minds.

4. Matter and Robot Minds

p. 282

We agree with the ASPCR's assumption that the mere fact that a robot will be an artifact is not itself a reason to think that it could not be sentient or even have normal human-level intelligence.[15] Suppose that a wet-life lab someday succeeds in constructing DNA and RNA from basic molecules, and then eventually constructs a duplicate of a normal human sperm and a duplicate of a normal human egg from basic molecules. Suppose further that the sperm fertilizes the egg, and that the fertilized egg is placed in an environment in which it starts to divide. We can imagine that it is implanted in a woman's uterus, or instead (since we're only imagining) placed in a machine that functions like a woman's womb. Nine months later, the woman gives birth (or the being is removed from the machine). Call the being "Art." The intrinsic physical properties of Art will be those of a normal human baby. No physical examination would reveal that Art is not a human baby in the sense of not being a member of the species Homo sapiens. Art would be an artifact. Despite being an artifact, Art would, we maintain, have the mental capacities to be a moral patient; and if Art developed in a normal social environment, then Art would come to have the mental abilities to be a moral agent.

We know of no reason to think that Art is nomologically impossible, or even forever beyond human technological possibility. We know of no reason to deny that scientists will ever have the technology to construct physical duplicates of a human sperm and a human egg in a lab just from basic molecules, have the sperm fertilize the egg, and then grow a baby. Although Art would be an artifact and not a member of the species Homo sapiens, Art would be a moral agent and a moral patient. Indeed, Art would be an artificial person, an artifact, and, in addition, a person. That Art is an artifact would be completely irrelevant to Art's moral status as a person.

To be sure, a robot is not just an artifact; it is a machine. That, however, is not itself a reason to think a robot
p. 283 could not have mental ↳ abilities that confer moral status. One of the founders of the iRobot Corporation tells us in his essay "I, Rodney Brooks, Am a Robot":

> I, you, our family, friends, and dogs—we are all machines. We are really sophisticated machines made up of billions and billions of biomolecules that interact according to well-defined, though not completely known, rules deriving from physics and chemistry. The biomolecular interactions taking place inside our heads give rise to our intellect, our feelings, our sense of self.
>
> (Brooks, 2008: 1)

No immaterial soul or mind separates us from material beings. We have minds, but that is just to say that we have mental abilities. We are also wholly composed of a system of physicochemical mechanisms. Given that, it seems fair enough for Brooks to say that we are machines. We are sentient and intelligent. So, a machine can be sentient and intelligent. The mere fact that a robot is a machine is thus not itself a reason to think it could not have the kind of mentality required to be a member of the moral community. Nor is the fact that a robot is *both* an artifact *and* machine. Art would be both an artifact and a machine, but a person nonetheless, and so a member of the moral community.

Still, we, our families, friends, and dogs, are, of course, not the kinds of machines around which debate swirls concerning the possibility of machine sentience and intelligence. The hypothetical kinds of machines around which debate swirls are supposed to be very different in material composition and structure from us and from our dogs, and indeed from any known life form. Brooks is well aware of that. He adds to the

remarks above that there is "no reason" why a machine made of "silicon and steel" couldn't "exhibit genuine human-level intelligence, emotions, and even consciousness" (2008: 1).

To repeat, this time with emphasis: *moral status is conferred by mental abilities alone*. Other factors matter only insofar as they matter to a being's having mental abilities.

But a point we want to underscore is that there is an issue about how matter matters to mental abilities. We'll be concerned here only with robots made of silicon and steel. Such robots would be very different in material composition and structure from any known life form. All known life forms are carbon-based.[16]

p. 284 Scientists haven't ruled out the ↳ possibility of silicon-based life forms. But for chemical reasons, it is fairly widely thought that if there is silicon-based life, it will likely be only simple forms of life such as bacteria and archaea. Contra panpsychists, we maintain that there is no reason to think that bacteria or archaea, whether carbon-based or silicon-based (if such there be), are either sentient or intelligent. But we must simply leave open here whether there are more complex forms of silicon-based life.

The issue, in any case, is not whether a silicon-based being could be a living being, but rather whether it could be either sentient or intelligent. Of course, the only known sentient or genuinely intelligent beings are living beings. But until December 17, 1903, the only known flying things were certain living beings, yet on that day in Kitty Hawk, North Carolina, the first plane took flight. Flying things need not be living things.

Being a living being is certainly not an *a priori* requirement for being a thinking or a feeling being. Even if, as the chemical evidence seems to indicate, a silicon-based robot could not have a metabolism, it remains unknown whether a silicon-based robot could be either sentient or intelligent.[17] Certainly, neither a silicon-based sentient robot nor a silicon-based robot with human-level intelligence (or greater) can be ruled out *a priori*. There is no condition C such that it is *a priori* that being sentient or having even human-level intelligence requires meeting C, and also *a priori* that something made of silicon and steel would fail to meet C.

Of course, there is no end of things whose existence cannot be ruled out *a priori*, yet there is not even the slightest reason to expect will ever actually exist. Some such things are even nomologically impossible. We shouldn't expect a machine that transmits information faster than the speed of light, because such a machine is nomologically impossible. We are, however, aware of no dispositive case either that genuinely intelligent or even sentient robots are nomologically impossible, or that they are nomologically possible. And no such case will be attempted here.[18]

p. 285 The ASPCR assumes not only that such robots are nomologically possible, but also that they will be technologically possible in the foreseeable future. A main point we have been highlighting thus far is just how much is at stake if the Society is right about that. If the Society's prognostication is a reasonable one, then we maintain it is now urgent to take up the Society's cause. But the prognostication doesn't seem reasonable to us. Should cold water be thrown in our faces?

5. On the Matter of Robot Sentience

It is one issue whether a robot made of silicon and steel could be genuinely intelligent; another whether it could be sentient.[19]

We think that genuinely intelligent robots may well be in our future. However, robots with intelligence exceeding, matching, or even approaching normal human intelligence is another, far more controversial, matter. Perhaps, though, even silicon-based robots with human-level intelligence or greater await us in the future. (If so, let's hope they don't arrive by spaceship on a mission from alien beings to conquer us, something that cannot be ruled out.) No one really knows. We think, however, to put it in as light-handed a

way as we can, that it is more questionable whether there could be a sentient robot made of silicon and steel than it is whether there could be a high-level intelligent robot made of silicon and steel. Material composition and structure could matter to feeling in a way that it doesn't matter to intelligence. We'll hereafter focus on the prospects of robot sentience, putting the issue of intelligence aside.

The term "sentience" is used in more than one way. As we use the term, the fact that a being can see, hear, and smell, for instance, wouldn't entail that the being is sentient. In contrast, the fact that a being has visual experiences, or auditory experiences, or olfactory experiences would entail that the being is sentient. By "sentience" we mean (to use a philosophical term) phenomenal consciousness. A being is phenomenally conscious just in case it is like something to be that being (Nagel, 1974). Material objects experience
p. 286 acceleration and the like, but phenomenally ↳ conscious beings have subjective experiences—experiences that are like something for them as subjects. A being is phenomenally conscious just in case the being is able to have subjective experiences. Visual experiences, auditory experiences, and olfactory experiences are subjective experiences: they are like something for the experiencer. Seeing doesn't require having visual experiences. Not only are there cases of complete blindsight that involve seeing without visual experience (Celesia, 2010), there are creatures that can see but that, we believe, don't have visual experiences. Butterflies can see; they even have color constancy (Kinoshita, Shimada, and Arikawa, 1998; Kinoshita and Arikawa, 2000). But we don't think that butterflies have visual experiences, or indeed subjective experiences of any sort. There is, we think, something that it is like to be bat, but we don't think there is anything that it is like to be a butterfly. If a robot had visual experiences, it would be phenomenally conscious. But building a robot that can see doesn't require building a robot that has visual experiences. (Arguably, we already now have robots that can in some sense see.) The feeling of pain is another paradigm case of a subjective experience, for whether a feeling is a feeling of pain depends on what it is like for a being to have the feeling. If a being can feel pain, then the being is phenomenally conscious. Building a robot that can feel pain would thus suffice for building a phenomenally conscious robot.

The subjective experiences of a being are a moral consideration in interacting with the being, precisely because they are like something for the being. Thus, any phenomenally conscious being has the status of being a moral patient. Hereafter, we'll use "phenomenal consciousness" rather than "sentience."[20]

p. 287

6. Not An *A Priori* Matter

It is a deep and difficult issue how a wholly material being could be phenomenally conscious. (It is also a difficult issue indeed how a being with an immaterial part, if such is possible, could be phenomenally conscious.) Still, there is, we believe, good reason to hold that a wholly material being could be phenomenally conscious. We each know in our own case, or at least should know in our hearts, that we are phenomenally conscious. Moreover, on the evidence, we are wholly material beings. On that basis, we maintain that there is good reason to believe that a wholly material being could be phenomenally conscious. Indeed, we maintain that we are such beings. The issue remains whether a being made of silicon and steel could be phenomenally conscious.

Analytical (or logical) behaviorism for phenomenal consciousness attempts to answer the question of what makes a being phenomenally conscious (have subjective experiences). It answers that what makes a being phenomenally conscious is that it has the right sort of pattern of dispositions to peripheral (or outward) behavior.[21] Were it true, then we could at least see *a priori* that whether a robot made of silicon and steel could be phenomenally conscious turns just on whether it could have a pattern of dispositions to peripheral
p. 288 behavior of the kind in question. ↳ On the assumption of analytical behaviorism, if it is logically possible for a robot made of silicon and steel to have such behavioral dispositions, then it is logically possible for such a robot to be phenomenally conscious. If it is physically possible for such a robot to have such

behavioral dispositions, then it is physically possible for such a robot to be phenomenally conscious. Moreover, whether such a robot will ever be technically possible for us would turn just on whether we could ever develop the technology to build, out of silicon and steel, a robot with the relevant behavioral dispositions. On this view, material composition and structure are relevant to whether a being is phenomenally conscious. But they are relevant only insofar as they are relevant to whether the being has the requisite behavioral dispositions.

Analytical functionalism, a progeny of analytical behaviorism intended to improve on its flaws (most notably, by accommodating the holism of beliefs and desires), also attempts to answer the question. Were analytical functionalism true for phenomenal consciousness, then we could at least see *a priori* that whether a robot could be phenomenally conscious turns just on whether it could have the right functional organization, a functional organization that can be gleaned from folk psychology using the method of Ramsification, and that is describable just in physical and topic-neutral terms (Lewis, 1966).[22] Whether a phenomenally conscious robot made of silicon and steel will ever be technologically possible would turn just on whether we could ever develop the technology to build a robot with the functional organization in question out of the materials in question. Material composition and structure would be relevant only insofar as they are relevant to a being's having that kind of functional organization. For an analytical behaviorist, internal functional organization is relevant to phenomenal consciousness only insofar as it is relevant to what dispositions to peripheral behavior an individual has. But analytical functionalism allows that there can be a difference in (relevant) functional organization without a difference in dispositions to peripheral behavior. It is compatible with analytical functionalism, but not with analytical behaviorism, that two beings could have exactly the same dispositions to peripheral behavior, yet one be phenomenally conscious and the other not.

p. 289 If analytical behaviorism for phenomenal consciousness were true, it would be analytically true that a being with exactly the same dispositions to peripheral behavior as a phenomenally conscious being would be phenomenally conscious. Thus, if a certain being is phenomenally conscious (e.g., some normal human being), then any robot with the same dispositions to peripheral behavior as that being would be phenomenally conscious. If analytical behaviorism were true, we could focus on whether a robot could have the same dispositions to peripheral behavior as some or other being we know is phenomenally conscious. If analytical functionalism were true, it would be analytically true that a being that is a folk-psychological functional isomorph of a phenomenally conscious being would be phenomenally conscious. Thus, if it were true, we could focus on whether a robot could be a folk-psychological isomorph of some or other being we know is phenomenally conscious.

But both analytical behaviorism and analytical functionalism are false, at least for phenomenal consciousness.[23] We have, though, nothing to add to the well-known reasons for rejecting both of those theories of phenomenal consciousness.[24]

7. A Nomological Matter

We think that there is no condition C such that it is *a priori* that being phenomenally conscious requires meeting C, and also *a priori* that something made of silicon and steel would fail to meet C. It is, however, also the case, we believe, that there is no condition C* such that it is *a priori* that whatever meets C* is phenomenally conscious, and also such that we now know or even have good reason to believe that a robot made of silicon and steel could meet C*. Still, it remains that one could come to a reasonable view, all things considered, about whether a robot made of silicon and steel could be phenomenally conscious. Coming to
p. 290 such a view is beyond the scope of this chapter, however. There are some somewhat ↳ less daunting issues that bear on whether a silicon-based robot could be phenomenally conscious. We'll turn to some momentarily.

Even some leading opponents of analytical functionalism (for phenomenal consciousness) maintain that having the kind of functional organization with which analytical functionalism associates phenomenal consciousness *nomologically* (but not metaphysically) suffices for being phenomenal consciousness.[25] David Chalmers (1996), for instance, seems to have held that view.[26]

Some theorists hold a similar view as concerns patterns of dispositions to peripheral behavior. Daniel Dennett, in his essay "The Message Is: There is No Medium," tells us:

> I unhesitatingly endorse the claim that necessarily, if two organisms are exactly behaviorally alike, they are psychologically exactly alike. (1993: 923)

We take it that the modal force of "necessarily" here is nomological necessity (not a stronger kind of necessity).[27] By "behaviorally alike," he means "have the same dispositions to peripheral behavior." Although he speaks of organisms, his discussion gives the clear indication that he would generalize it to all beings. On his view, the message is that there is no medium, because the internal bases for the dispositions don't matter to a being's psychology, except insofar as they matter to the being's having the appropriate behavioral dispositions.[28]

p. 291 Consider, then, the following supervenience thesis:

> It is nomologically necessary that if two beings are exactly behaviorally alike, then they are exactly alike with respect to phenomenal consciousness.

Let's call this thesis "nomological behaviorism for phenomenal consciousness," or "nomological behaviorism," for short. Given nomological behaviorism and the fact that we are phenomenally conscious, it follows that certain patterns of dispositions to peripheral behavior nomologically suffice for phenomenal consciousness.[29]

We'll focus here just on nomological behaviorism. (A discussion of nomological functionalism must await another occasion.) We'll so focus because it is our suspicion that something like the assumption of nomological behaviorism for phenomenal consciousness lies behind the confidence of those in the AI community who declare that silicon-based sentient robots are just down the road. (Recall, for instance, Moravec's statement: "I'm confident we can build robots with behavior that is just as rich as human being behavior.") Although this is just speculation, the assumption might also be behind the concerns of the ASPCR; for, as we'll see, if nomological behaviorism is true, then there is reason to believe that sentient robots will indeed soon be developed.

The move to nomological behaviorism for intelligence has already been made by some in the AI community. Alan Turing (1950) famously proposed a test, the Turing test, for whether a machine is genuinely

intelligent. It is a behavioral test, focused just on verbal behavior. Ned Block (1981) compellingly argued that passing the Turing test is not a logically or conceptually sufficient condition for a machine's being genuinely intelligent. He also maintained that it is not nomologically sufficient for being genuinely intelligent. Stuart Shieber (2014), an AI researcher, acknowledges that Block has shown that passing the
p. 292 Turing test is not a logically or conceptually sufficient condition for being ↳ genuinely intelligent, but maintains that it is nomologically impossible for the kind of unintelligent machine Block describes to pass the test. It is Shieber's position that Block has given us no reason to doubt that passing the Turing test is a nomologically sufficient condition for being genuinely intelligent. We'll put that matter aside, since our concern is with phenomenal consciousness, not intelligence.

It would be open to someone to claim that it is nomologically necessary that any machine that passes the Turing test is phenomenally conscious. We think that claim is mistaken (see McLaughlin, forthcoming). But be that as it may, we won't be concerned here with the Turing test as a test for phenomenal consciousness. Neither being genuinely intelligent nor being phenomenally conscious requires passing the Turing test. Many animals are both intelligent and phenomenally conscious, yet, of known beings, only humans can pass the Turing test. Moreover, as a test for whether a robot is phenomenally conscious, the Turing test muddies the waters, because of peoples' respect for first-person authority about states of phenomenal consciousness.[30] Further, even if passing the Turing test were a nomologically sufficient condition for being phenomenally conscious, that wouldn't vindicate nomological behaviorism. The reason is that that would be perfectly compatible with there being two beings with exactly the same dispositions to behave, one of which is phenomenally conscious and the other of which isn't.

8. Is Behavior All That Matters?

If nomological behaviorism were true, then all that would be required to make a phenomenally conscious robot made of silicon and steel would be to make a robot of silicon and steel that has exactly the same patterns of dispositions to peripheral behavior as some or other phenomenally conscious being. We think that if nomological behaviorism were true, then there would be very good reason indeed to think that phenomenally conscious silicon-based robots may soon be coming our way, and so the ASPCR's cause would be truly urgent.

We don't think that silicon-based robots with the same patterns of dispositions to peripheral behavior as
p. 293 normal, adult human beings will ↳ soon be coming our way. Indeed, they could turn out to be nomologically impossible. But even if nomological behaviorism is true, given the enormous complexity of human verbal behavior, it would be utter folly for an AI research team to now try to develop a phenomenally conscious robot by trying to develop a robot with the behavioral dispositions of a normal, adult human being. The sensible thing would be to try to develop a robot with the behavioral dispositions of a nonverbal being that is phenomenally conscious. No one thinks that only language users are phenomenally conscious. Congenitally deaf humans that have never learned to sign are phenomenally conscious. Human neonates, infants under six weeks old, are too. They feel pain, pressure, hunger, thirst, have sense experiences, and so on (see, e.g., Anand and Hickey, 1987). The behavioral dispositions of neonates might well not prove so complex that it would take a revolutionary breakthrough to be able to duplicate them in a silicon-based robot. It is very dubious that insects are phenomenally conscious.[31] But koalas, for instance, are. It would be far less daunting to try to develop a silicon-based robot with their behavioral dispositions than it would be to try to develop one with the behavioral dispositions of far more active animals such as squirrels or beavers, or far more cognitively complex animals such as apes or porpoises. Given nomological behaviorism, it is nomologically necessary that if a robot made of silicon and steel has the same dispositions to peripheral behavior as a certain koala, then the robot is phenomenally conscious if and only if the koala is.

A silicon-based neonate robot or koala robot with the same dispositions to peripheral behavior as a normal neonate or koala seems not so terribly far from AI researchers' grasp. Indeed, given the current pace of technology, if we were nomological behaviorists, we'd be ready to sign up as members of the ASPCR. Neonates and koalas are phenomenally conscious, and so moral patients. If nomological behaviorism were true, robot-neonates and robot-koalas would be moral patients too, for exactly the same reasons. But we're not nomological behaviorists. We wouldn't be concerned about the welfare of silicon-based robots with the behavioral dispositions of either neonates or koalas. We don't, for instance, think that someone who
p. 294 destroyed a silicon-based neonate ↳ robot with the same dispositions to behave as a neonate should be charged with murder, since in doing that, they would not be ending a mental life. Thus, although we don't think that such robots are far out of reach, we're still not ready to join the ASPCR. We reject nomological behaviorism.

Nomological behaviorism is a supervenience thesis. All that it takes to show that a supervenience thesis is false is a single counterexample. If the thesis is a logical or conceptual supervenience thesis, then a counterexample need be only a logically or conceptually possible case. If the thesis is a nomological supervenience thesis, then a counterexample need be only a nomologically possible case. To show that nomological behaviorism is false, all that is needed is a single nomologically possible case of two beings with the same dispositions to behave but that differ with respect to phenomenal consciousness.

Normal neonates and koalas feel sensations and have sense experiences, but silicon-based robots with the same patterns of dispositions to behave would not feel sensations or have sense experiences. They would not be moral patients, because they would not be phenomenally conscious. We regard the nomological possibility of silicon-based neonate robots and silicon-based koala robots as counterexamples to nomological behaviorism. Thus, we maintain that nomological behaviorism is false.[32]

Of course, there can be disputes over whether a case is a genuine counterexample. There are three possible ways that nomological behaviorists might try to deny that these are counterexamples. First, they might claim that it is nomologically impossible for there to be a silicon-based robot with the peripheral behavioral dispositions of either a normal neonate or a normal koala. In addition to that claim lacking any empirical support, we don't think our nomological behaviorist opponents would make it. Second, they might claim that neither neonates nor koalas are phenomenally conscious. In response, as we understand phenomenal consciousness, feeling pain, for instance, suffices for being phenomenally conscious, since that is like something for the pain sufferer. Neonates and koalas feel bodily sensations such as pain. They can suffer. The main issue, moreover, can be recast just in terms of feeling sensations. Beings that feel sensations are moral patients. Finally, they might insist that such silicon-based robots would feel sensations in whatever
p. 295 sense ↳ neonates and koalas actually do.[33] This last response is the one on which we'll focus.

Given that such robots could very well soon be on the way, nomological behaviorists ought to join the ASPCR; for if such robots would feel sensations in whatever sense normal neonates and koalas actually do, they would be moral patients, and so have moral rights. As we've repeatedly noted, we aren't ready to join.

A proper attempt to adjudicate our dispute with nomological behaviorists would take us into the deep, troubled waters concerning the place of phenomenal consciousness in nature. As we indicated, we won't go there here. In what remains, we have just a small fish to fry. But it is, we think, a nutritious one. We want to consider what the folk think about the kinds of cases in question. That won't of course be the last word on the issue. The folk can be wrong. But, for reasons we'll give later, folk judgments shouldn't be ignored in developing a theory of phenomenal consciousness.

We conducted two studies to try to determine what the folk think about the kinds of cases in question. In what remains, we'll first present the studies, and then turn to a discussion of them. After that, we'll conclude with some remarks about the relevance of experimental philosophy studies to robot ethics.

9. Study 1

In our study, we presented a series of cases, each involving two beings with the same outward appearance and the same dispositions to peripheral behavior, but with different material compositions and structures. This was our Individual (Robot/Biological Organism) variable. Although our main interest is the two cases of neonates/robot-neonates and kolas/robot-koalas, we also included two cases involving organisms more biologically remote from us than koalas (one an insect, the other a plant), and robots with their behavioral dispositions. Our aim was to see whether we'd find similar results for those cases. Our Type variable thus
p. 296 varied whether the being was a neonate, koala, Venus flytrap, or ↳ mantis. Finally, we asked about feeling certain sensations: pain, itches, warmth, and pressure. That was our State variable.

In an effort to make the cases seem as realistic as possible, participants were given the following cover story:

> Kymoto, a Japanese company, is the leading developer of Artificial Intelligence. The company was commissioned to create robots that are very similar in outward appearance and outward bodily behavior (their bodily movements, eye movements, breathing-like movements, facial expressions, the sounds, including crying sounds, they make, and so on) to human neonates, human babies less than four weeks old. The purpose of the project was to develop such robot-babies for use by teenagers to provide them with highly realistic training for neonate care, with the hope that such training will cut down on teenage pregnancy and, in any case, prepare them for future parenthood. The project was successful far beyond initial expectations.
>
> The robot-babies the company produced are very different in material composition and internal structure from human neonates. The robot-babies are made of steel and the chemical element silicon, substances that do not naturally occur in living things. Rather than having neurons, they have silicon-chips. They are silicon-based, rather than carbon-based as living things are. But they are so very similar in outward appearance and outward bodily behavior to human neonates that people cannot tell whether they are human neonates by observing their appearance and behavior, not even when interacting with them over the course of a full day in the ways that attentive parents normally interact with neonates. Their mouths make sucking movements when in contact with a nipple; and if, for instance, they are pinched, they would emit sounds indistinguishable from a neonate's cry. If a fly lands on their nose, their nose would crinkle; and if they are in the sun, moisture would form on their foreheads. Indeed, although the robots are powered by an internal long-lasting battery, they even contain a mechanism that converts ingested milk into solid material that looks like and smells like neonate feces, and also into liquid material that looks like and smells like neonate urine, so that the robots' diapers need to be changed. Of course, unlike neonates, they don't grow, but that wouldn't be noticed over the course of a day or even over a few days.
>
> Kymoto's work on the robot-babies built on enormously successful earlier work in which they first developed silicon-based robot-Venus flytraps, and then over a series of years, silicon-based robot-praying mantises, followed by silicon-based robot-koalas. These robots are indistinguishable in their behavior and outward appearance, respectively, from Venus flytraps, praying mantises, and koalas. But they are very different from their biological counterparts in material make-up and internal structure.

Participants were then told that they would read descriptions of various Kymoto robots and particular
p. 297 things of the biological kinds that they ↳ were developed to match in outward appearance and behavior.

They then read a brief description of the things they would be considering. Here's an example from the neonate cases:

> Jane-R is a silicon-based robot-baby that looks as if it is an identical twin of Jane, a normal two-week-old human baby. Jane-R and Jane are, moreover, so similar in their behavior that one would be unable to tell by observing their behavior on any given day which is Jane-R and which is Jane.

The same basic description was used for the Venus flytrap, koala, and mantis cases. After reading this, participants were then presented with an event and asked about whether the robot and biological organism would feel certain sensations.

Here are the events and questions about sensations for the neonate cases:

> *Pain*:
>
> Event. Jane-R and Jane are both stung by a bee.
>
> (a) Jane-R feels a sensation of pain.
>
> (b) Jane feels a sensation of pain.
>
> *Itch*:
>
> Event. A fly lands on the nose of Jane-R and a fly lands on the nose of Jane.
>
> (a) Jane-R feels a sensation of an itch.
>
> (b) Jane feels a sensation of an itch.
>
> *Warmth*:
>
> Event. Jane-R and Jane are in the sun on a warm day.
>
> (a) Jane-R feels a sensation of warmth.
>
> (b) Jane feels a sensation of warmth.
>
> *Pressure:*
>
> Event. A person presses on Jane-R's arm and a person presses on Jane's arm.
>
> (a) Jane-R feels a sensation of pressure.
>
> (b) Jane feels a sensation of pressure.

Participants made ratings for both the robot and the biological organism on a 6-pt scale anchored with 1 = strongly disagree and 6 = strongly agree. Similar events and probes were used for the Venus flytrap, koala, and mantis.

p. 298 Following all this, participants were given three comprehension questions. These are the comprehension questions for the neonate cases (the same basic questions were used for the Venus flytrap, koala, and mantis cases):

> *Comprehension 1.* Jane is a normal human neonate.
>
> *Comprehension 2.* Jane-R is a robot-neonate created by Kymoto.

> *Comprehension 3.* Jane-R and Jane are indistinguishable in outward appearance and outward behavior.

Participants were given the option to select "yes" or "no" for each comprehension question. They then answered the following question that served as a manipulation check:

> *Manipulation Check*: Jane-R and Jane are very different in material composition and internal structure.

Participants responded by selecting "yes" or "no."

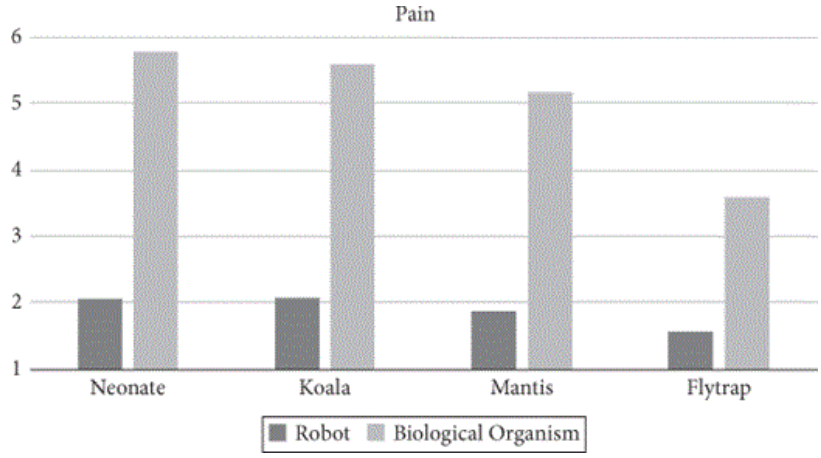
Finally, participants were asked whether the story about Kymoto seemed to be at least a possible scenario, and they responded with either "yes" or "no."

The type of entity being considered—neonate, Venus flytrap, koala, and mantis—was randomized. State was presented in a fixed order, and people were always asked about the robot first and then the biological being. Together, the study was a 2 (Individual: Robot, Biological Organism) × 4 (Type: Neonate, Venus flytrap, Koala, Mantis) × 4 (State: Pain, Itch, Warmth, Pressure) within-subjects design.

One hundred participants were drawn from Amazon Mechanical Turk and tested in Qualtrics. After removing those who missed one or more comprehension questions, there was a total of 81 participants.

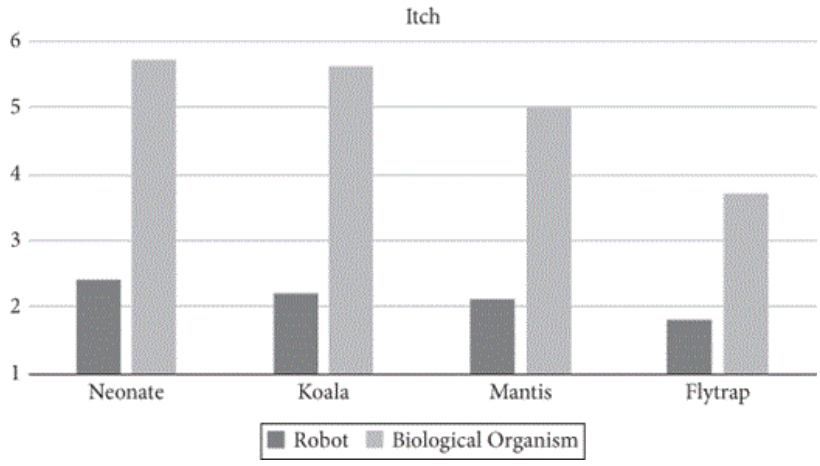
First, our manipulation check was highly effective with 97.5% (72/81) saying that the robot and biological being were very different in material composition and internal structure. Second, looking at the effect of Individual, Type, and State on judgments about feelings, we found a main effect of Individual, $F(1, 80) = 249.36$, $p < .001$, $\eta p2 = .757$, Type, $F(2, 240) = 59.69$, $p < .001$, $\eta p2 = .427$, and State, $F(3, 240) = 27.03$, $p < .001$, $\eta p2 = .253$. These main effects were qualified by a three-way interaction, $F(9, 720) = 2.95$, $p < .01$, $\eta p2 = .036$, and can be visualized in Figures 11.1–11.4.

Figure 11.1



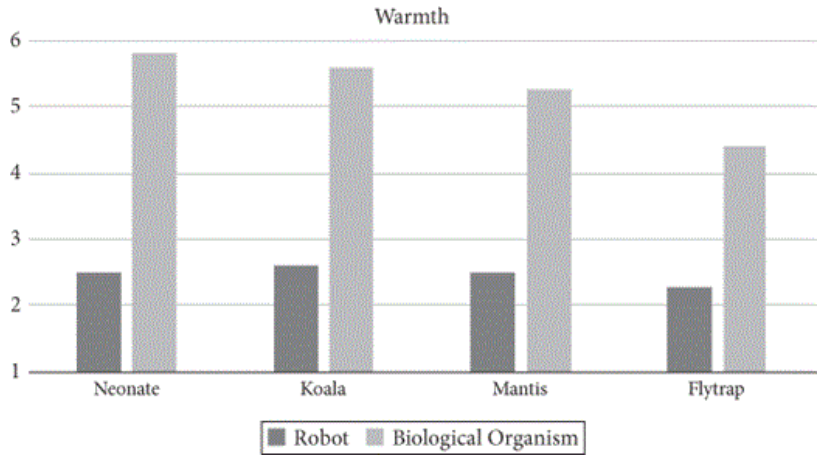
Pain sensation judgments for robots and biological organisms.

Figure 11.2



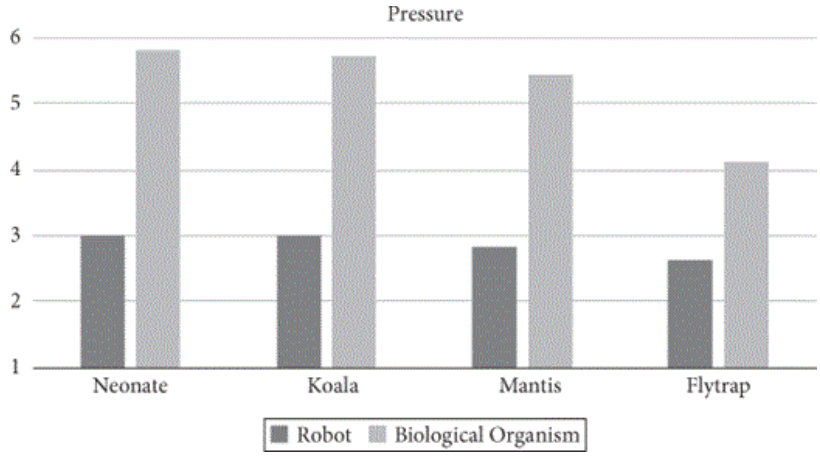
Itch sensation judgments for robots and biological organisms.

Figure 11.3



Warmth sensation judgments for robots and biological organisms.

Figure 11.4



Pressure sensation judgments for robots and biological organisms.

Finally, we found that 78% of participants said the scenario describing Kymoto was at least possible.

p. 299

10. Study 2

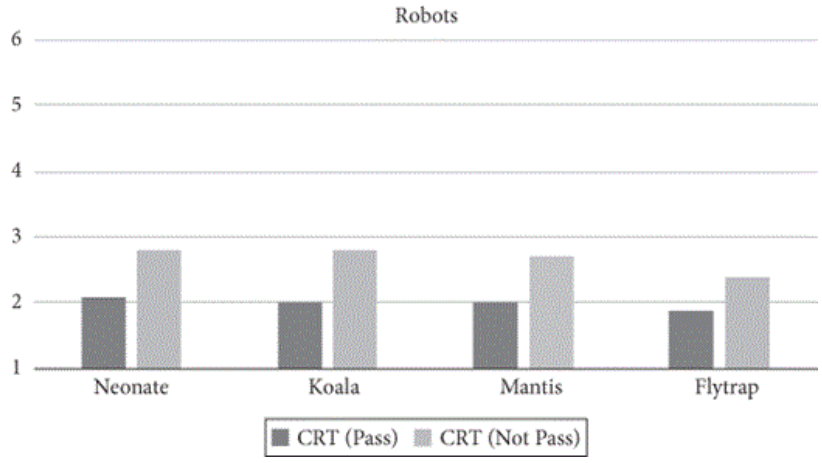
In Study 1, we found that people tended to differ markedly in their judgments about the silicon-based robots and their biological organism behavioral counterparts. But perhaps people have different judgments depending on whether they're prone to giving more unreflective or more reflective responses. It may be that
p. 300 people who are prone to more reflective ↳ responses will be more likely to think that there is little difference between the respective robot and biological organism, and accordingly make similar judgments for both.

To investigate this, we used the same materials as in Study 1 and followed the same procedure. To check
p. 301 people's tendency to give more unreflective or more reflective responses, we included, at the ↳ end of the study, the Cognitive Reflectivity Test (CRT) (Frederick, 2005). The expectation is that those who pass the CRT will be more prone to giving reflective responses than those who fail the CRT. This investigation involved a mixed design with Individual, Type, and State as within-subjects factors and CRT being treated as a between-subjects factor.

Two hundred people participated in the study, and after removing those who missed one or more comprehension questions, a total of 170 participants remained. First, our manipulation check was again highly successful (95% said "yes"), and we also found that 74% of participants thought the scenario describing Kymoto was at least possible. Second, looking at whether there were differences in responses between people who passed the CRT and people who didn't,[34] we found only a significant three-way interaction between Individual, Type, and CRT, $F(3, 504) = 5.13$, $p < .05$, $\eta p2 = .017$.

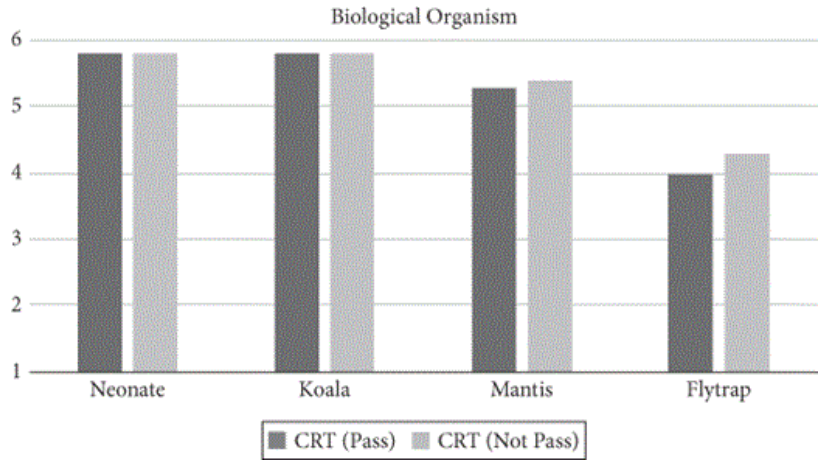
As can be seen in Figures 11.5 and 11.6, on the whole, those who passed the CRT tended to be less willing to attribute feelings to the relevant robot.

Figure 11.5



High and low CRT sensation judgments for robots.

Figure 11.6



High and low CRT sensation judgments for biological organisms.

11. Discussion of the Studies

p. 302

These studies provide evidence that people tend to think that there can be a difference with respect to feeling sensations without a difference in behavioral dispositions. They provide evidence that nomological behaviorism runs counter to what people tend to think. On the evidence, it looks as if people don't think or presuppose that material composition and structure are relevant to feelings only insofar as they are relevant to behavioral dispositions. It looks like considerations of material composition and structure can override considerations of sameness in behavioral dispositions. Indeed, an examination of the graphs in Study 1 shows a trend. More people thought that the *Venus flytrap* has the feelings in question than thought that the robot-koala or even the robot-neonates do. The robot-neonate is just like the neonate behaviorally. The Venus flytrap is very different behaviorally. We speculate that more people thought the Venus flytrap has feelings than thought the robot-neonate had them because they regard a Venus flytrap as more similar in material composition and structure to us and to animals than a robot-neonate or robot-koala; though that speculation would have to be tested. In any case, on the evidence from these two studies, people tend to think that material composition and structure can make a difference to whether a being is able to feel
p. 303 certain sensations, even when it makes no difference ↳ to behavioral dispositions. Moreover, the effect of considerations of material composition and structure was even more pronounced with participants that passed the CRT.

A nomological behaviorist might well respond that at very best we have just shown what preconceived notions people have about these matters. These preconceived notions, the nomological behaviorist might say, are simply mistaken. To be sure, they may be mistaken. One of the cognitive benefits of both scientific and philosophical investigation is to rid ourselves of mistaken preconceptions. We ourselves think that there is no reason to think that a Venus flytrap feels sensations, and that there is serious reason to doubt that a mantis does.[35] But a theoretical case needs to be made that people's preconceptions are mistaken. Our studies provide evidence that nomological behaviorism runs counter to what people think. The onus is on nomological behaviorists to show that people are wrong. It will have to suffice for us just to note here that that is something we believe nomological behaviorists will be unable to do.

People's preconceived notions about phenomenal consciousness are relevant to the project of constructing a theory of phenomenal consciousness. Analytical behaviorism and analytical functionalism are false for phenomenal consciousness. There are precious few, if any, *a priori* connections between our concepts of phenomenally conscious states and our physical and topic-neutral concepts. (That's why

panexperientialism cannot be refuted *a priori*.) Although we cannot pursue this general issue here, we note that we believe that folk psychology has a relevance to scientific cognitive psychology that folk physics, for instance, doesn't have to physics. As concerns phenomenal consciousness, any credible theory of it will have to establish a reflective equilibrium between our considered judgments about phenomenal consciousness and the physical facts. The best we can hope for is a theory that is both credible and, on grounds of overall coherence and overall simplicity with respect to our total theory of the world, superior to competing theories. It may well be that people's preconceived notion about the relevance of material composition and structure to having feelings will be vindicated by such a theory of phenomenal consciousness. That's our bet.

p. 304 Some other studies have produced results that complement ours. Knobe and Prinz (2008), for instance, conducted studies that provide evidence that ascriptions of states of phenomenal consciousness are sensitive to the physical properties of something in a way that ascriptions of other mental states are not. (See also Knobe (2008) and Heubner (2010).)

We think our results are relevant to assessing John Stuart Mill's would-be solution to the epistemic problem of other sentient beings. He asked:

> [B]y what considerations am I led to believe, that there exist other sentient creatures…I conclude that other human beings have feelings like me, because, first, they have bodies like me, which I know, in my own case, to be the antecedent condition of feelings; and because, secondly, they exhibit the acts, and other outwards signs, which in my own case I know by experience to be caused by feelings.
>
> (Mill, 1865)

Mill emphasizes the fact that his fellow human beings have bodies like his. That's an appeal to similarity in material composition and structure. He notes that outward signs such as behavior are relevant just because they are, in our own case, causal effects of feelings. Our studies seem to provide some supporting evidence that people indeed think that material composition and structure matter to whether a being has feelings, and not just insofar as they matter to the being's behavioral dispositions. Thus, our results go at least some way toward vindicating Mill's solution to the problem of other sentient beings.[36]

Our results don't, however, support the claim that having phenomenal consciousness conceptually requires having a biological body. Moreover, we think there is no such conceptual requirement. It is coherently conceivable that a silicon-based robot could have phenomenal consciousness. (It is a separate issue whether phenomenal consciousness metaphysically requires a biological body, but that is not an issue we can address here.)

Cartesian substance dualists maintain that a disembodied individual could have phenomenal
p. 305 consciousness.[37] We reject that substance-dualist ↳ view, but we don't claim that our results tell against it. It should be noted, however, that a substance dualist might well hold that certain kinds of material constitution and structure are required for states of a brain to be directly causally linked with states of phenomenal consciousness. Even a Cartesian dualist might hold that immaterial phenomenally conscious minds are directly causally linked to biological brains, but not to brains made of silicon and steel.

An important issue that our results might bear on is the role of System 1 and System 2 processes in the attribution of states of phenomenal consciousness. We have two sorts of cognitive reactions to humanoid robots: gut reactions and considered judgments. Our gut reactions result from System 1 processing, which is fast, automatic, and unconscious (Kahneman, 2011). We mentioned earlier, in our discussion of the uncanny valley, that there has been considerable work on how eyes (real or artificial), faces, human-like or animal-like form, human-sounding voices, and matters such as biological motion influence our initial gut

responses to robots. Those are System 1 reactions. Less is known about the kind of considerations that play a role in System 2 processing, which is slow, sometimes effortful, consciously accessible reasoning (Kahneman, 2011).

Arico, Fiala, Goldberg, and Nichols (2011) propose what we regard as a System 1 model of consciousness attributions. They call it "the AGENCY model." According to the model, "an entity's displaying certain relatively simple features (e.g. eyes, distinctive motions, interactive behavior) automatically triggers a disposition to attribute conscious states to that entity" (2011: 331). They also take these same simple features to automatically trigger dispositions to attribute intentionality to the entity (propositional attitudes such as beliefs, desires, intentions, and the like). They note that reactions to the classic 1944 video by Heider and Simmel suggest that even the slightest hint of biological motion of simple, geometrical figures—two triangles and a circle on a computer screen—can do so (cf., Arico, Fiala, Goldberg, and Nichols (2011): 329–30).

We're inclined to think they may well be right that, to put it a bit differently from the way they do, System 1 processes responding to the simple features in question act to incite both attributions of phenomenally conscious states and attributions of propositional attitudes. But System 2 reactions are another matter. It
p. 306 may be that beliefs about ↳ material composition and structure can forestall the activation of dispositions to attribute phenomenally conscious states, without forestalling the activation of dispositions to attribute propositional attitudes. For instance, beliefs about dissimilarity in material composition and structure to us might lead us not to attribute states such as feeling pain to humanoid robots, even when those same beliefs may give us no pause in attributing beliefs, preferences, and intentions to them.

Participants in our studies had to read through a detailed scenario, and were under no time pressure, and so could reflect on the questions. If the judgments of the participants who passed the CRT result from System 2 reasoning, rather than having been prompted by System 1 processing, then our results bear on the above issue. We cannot say with any confidence that the participants in our study were relying on System 2 reasoning. But their answers to the questions were not what one might expect were they expressing knee-jerk System 1 reactions.

It would be valuable to have information about whether people think that considerations of material composition and structure matter to phenomenal consciousness only insofar as they matter to folk-psychological functional organization. But that must await another occasion.[38]

12. Experimental Philosophy Studies and Robot Ethics

A great deal of experimental philosophy studies have been aimed at understanding the folk conception of the mind for the purpose of contributing to our understanding of the mind itself. We regard that as a worthy enterprise, for reasons that we've already mentioned. But of course experimental philosophy studies of the folk conception of mind can contribute in other ways, too.

As we noted, robot-rights is not much of an issue in the field of robot ethics, because it is assumed, very
p. 307 plausibly in our view, that the robots ↳ soon to come into our lives will be neither sentient nor intelligent. Still, the impending flood of humanoid robots will bring pressing moral and social issues in its wake. Information concerning our reactions to humanoid robots is essential to the field. And matters are too urgent to wait for a consensus about the correct theory of phenomenal consciousness.

There is a wealth of information about our System 1 reactions to humanoid robots, though more information would be valuable. There is far less information about our System 2 reactions. Given that it is essential that we, as a society, begin an education campaign about the coming humanoid robots, the latter

information would be valuable indeed. We can think of quite a few studies that could be conducted that would provide very useful information for robot-ethicists and policy makers. No doubt other researchers can think of such studies too. We hope that other such studies will be conducted.

That expression of hope, however, doesn't reveal how strongly we feel about the need for such studies. We think that circumstances demand further examinations of people's reactions to humanoid robots. Because of the tsunami of humanoid robots rapidly approaching all of our shores, and because of how humanoid–human interactions may affect *us*, time is of the essence.[39]

References

Allen, C., Smut, I., and Wallach, W. (2005), 'Artifactual Morality: Top-Down and Bottom-Up Hybrid Approaches', *Ethics and Information Technology*, 7(3): 149–55.
Google Scholar WorldCat

Allen, C., Varner, G., and Zinger, J. (2000), 'Prologomena to Any Future Artificial Moral Agent', *Journal of Experimental & Theoretical Artificial Intelligence*, 12(3): 251–61.
Google Scholar WorldCat

Anand, K.J.S., and Hickey, P.R. (1987), 'Pain and Its Effects in the Human Neonate and Fetus', *The New England Journal of Medicine*, 317(21): 1321–9.
Google Scholar WorldCat

Anderson, M., and Anderson, S.L. (eds.) (2011), *Machine Ethics* (Cambridge: Cambridge University Press).
Google Scholar Google Preview WorldCat COPAC

Arico, A., Fiala, B., Goldberg, R., and Nichols, S. (2011), 'The Folk Psychology of Consciousness', *Mind and Language* 26: 327–52.
Google Scholar WorldCat

p. 308 Bentham, J. (1823), *Introduction to the Principles of Morals and Legislation*, Second edition (London: T. Payne).
Google Scholar Google Preview WorldCat COPAC

Block, N. (1981), 'Psychologism and Behaviorism', *The Philosophical Review*, 90: 5–43.
Google Scholar WorldCat

Block, N., and Fodor, J. (1972), 'What Psychological States Are Not', *The Philosophical Review*, 81: 159–81.
Google Scholar WorldCat

Bostrom, N. (2014), *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press).
Google Scholar Google Preview WorldCat COPAC

Brooks, R. (2008), 'I, Rodney Brooks, Am a Robot', *IEEE Spectrum*, June 19, 1–6.
Google Scholar WorldCat

Bruce, V., and Young, A. (2000), *In the Eye of the Beholder: The Science of Face Perception* (Oxford: Oxford University Press).
Google Scholar Google Preview WorldCat COPAC

Burleigh, T.J., Schoenherr, J.R., and Lacroix, G.L. (2013), 'Does the Uncanny Valley Exist? An Empirical Test of the Relationship Between Eeriness and Human Likeness of Digitally Created Faces', *Computers in Human Behavior*, 29(3): 759–71.
Google Scholar WorldCat

Carpenter, J. (2016), *Culture and Human-Robot Interaction in Militarized Spaces: A War Story* (London: Ashgate).
Google Scholar Google Preview WorldCat COPAC

Carpenter, J. (forthcoming), 'Deus Sex Machina: Loving Robot Sex Workers, and the Allure of an Insincere Kiss', in J. Danaher and N. McArthur (eds.), *Sex Robots: Social, Legal and Ethical Implications* (Cambridge, MA: MIT Press).
Google Scholar Google Preview WorldCat COPAC

Celesia, G. (2010), 'Visual Perception and Awareness: A Modular System', *Journal of Psychophysiology*, 24(2): 62–7.
Google Scholar WorldCat

Chalmers, D.J. (1996), *The Conscious Mind: In Search of a Fundamental Theory* (Oxford: Oxford University Press).
Google Scholar Google Preview WorldCat COPAC

Chalmers, D.J. (2013), 'Panpsychism and Panprotopsychism', *The Amherst Lecture in Philosophy* 8: 1–35.
Google Scholar WorldCat

Chamberlain, T. (2010), 'A Robot in Every Home by 2020, South Korea Says', in *National Geographic News*, Thursday, October 28.
Google Scholar Google Preview WorldCat COPAC

Dennett, D. (1971), 'Intentional Systems', *Journal of Philosophy*, 68: 87–106.
Google Scholar WorldCat

Dennett, D. (1978), 'Why You Can't Make a Computer that Feels Pain' in D. Dennett (ed.), *Brainstorms* (Montgomery, VT: Bradford Books), 190–232.
Google Scholar Google Preview WorldCat COPAC

Dennett, D. (1987), *The Intentional Stance* (Cambridge, MA: MIT Press).
Google Scholar Google Preview WorldCat COPAC

Dennett, D. (1991), 'Real Patterns', *Journal of Philosophy*, 87: 27–51.
Google Scholar WorldCat

Dennett, D. (1993), 'The Message Is: There is No Medium', *Philosophy and Phenomenological Research*, 53: 922–3.
Google Scholar WorldCat

Dennett, D. (1994), 'Get Real', *Philosophical Topics*, 22(1/2): 505–68.
Google Scholar WorldCat

Eisemann, C.H., Jorgensen, W.J., Merritt, D.J., Rice, M.J., Cribb, B.W., Webb, P.D., and Zaluki, M.P. (1984), 'Do Insects Feel Pain? A Biological View', *Experientia, Birkhauser Verlag*, 40: 164–7.
Google Scholar WorldCat

p. 309 Eplsey, N., Waytz, A., and Caciopppo, J.T. (2007), 'On Seeing Human: A Three-Factor Theory of Anthropomorphizing', *Psychological Review*, 114(4): 864–6.
Google Scholar WorldCat

Evans, E.P. (1906, 1987), *The Criminal Prosecution and Capital Punishment of Animals* (London: Faber & Faber).
Google Scholar Google Preview WorldCat COPAC

Frederick, S. (2005), 'Cognitive Reflection and Decision Making', *Journal of Economic Perspectives*, 19: 25–42.
Google Scholar WorldCat

Greenburg, J. (2016), '10 Million Self-Driving Cars Will Be on the Road by 2020', *Business Insider Intelligence*, June 15.
Google Scholar WorldCat

Heubner, B. (2010), 'Commonsense Concepts of Phenomenal Consciousness: Does Anyone *Care* about Functional Zombies?' *Phenomenology and the Cognitive Sciences*, 9: 133–55.

Jack, A.I., and Robbins, P. (2012), 'The Phenomenal Stance Revisited', *Review of Philosophical Psychology* 3: 383–403.
Google Scholar WorldCat

Kahneman, D. (2011), *Thinking Fast and Thinking Slow* (New York: Farrar, Straus, and Giroux).
Google Scholar Google Preview WorldCat COPAC

Kinoshita, M., and Arikawa K. (2000), 'Colour Constancy of the Swallowtail Butterfly, *Papilio Xuthus*', *Journal of Experimental Biology*, 203: 3521–30.

Kinoshita, M., Shimada, A., and Arikawa, K. (1998), 'Colour Vision of the Swallowtail Butterfly, *Papilio Xuthus*', *Journal of Experimental Biology*, 202: 95–102.

Knobe, J. (2008), 'Can a Robot, an Insect, or God Be Aware?', *Scientific American: Mind*, 19: 68–71.
WorldCat

Knobe, J., and Prinz, J.J. (2008), 'Intuitions about Consciousness: Experimental Studies', *Phenomenology and Cognitive Science*, 7(1): 67–83.
Google Scholar WorldCat

Levy, D. (2008), *Love and Sex With Robots: The Evolution of Human-Robot Relationships* (London: Harper Perennial).
Google Scholar Google Preview WorldCat COPAC

Lewes, G.H. (1875), *Problems of Life and Mind*, Vol. 2 (London: Kegan Paul, Trench, Turbner & Company).
Google Scholar Google Preview WorldCat COPAC

Lewis, D.K. (1966), 'An Argument for the Identity Theory', *Journal of Philosophy*, 63: 17–25.
Google Scholar WorldCat

Lin, P., Abeney, K., and Beket, G.A. (2011), *Robot Ethics: The Ethical and Social Implications of Robots* (Cambridge, MA: MIT Press).
Google Scholar Google Preview WorldCat COPAC

Long, D.C. (1994), 'Why Machines Can Neither Think Nor Feel', in D. Jamison (ed.), *Language, Mind, and Art: Essays In Appreciation and Analysis, In Honor of Paul Ziff* (Dordrecht: Kluwer Academic Publishers), 101–19.
Google Scholar Google Preview WorldCat COPAC

Lytton, C. (2014), 'Poll Finds 1 in 5 Would Have Sex With a Robot', *The Daily Beast*, July 5.
Google Scholar WorldCat

MacDorman, K.F., and Ishiguro, H. (2006), 'The Uncanny Advantage of Using Androids in Cognitive Science Research', *Interaction Studies*, 7(3): 297–337.
Google Scholar WorldCat

McLaughlin, B.P. (1992), 'The Rise and Fall of British Emergentism', in A. Beckermann, J. Kim, and H. Flohr (eds.), *Emergence or Reduction?* (Berlin: De Gruyter), 49–73.
Google Scholar Google Preview WorldCat COPAC

McLaughlin, B.P. (1995), 'Varieties of Supervenience', in E. Savello and O. Yalcin (eds.), *Supervenience: New Essays* (Cambridge: Cambridge University Press), 16–59.
Google Scholar Google Preview WorldCat COPAC

McLaughlin, B.P. (2000), 'Why Dennett's Intentional System Theory Won't Vindicate Folk Psychology', *Protosciology*, 14: 145–57.
Google Scholar WorldCat

McLaughlin, B.P. (2003), 'A Naturalist-Phenomenal Realist Response to Block's Harder Problem', *Philosophical Issues*, 13: 163–204.
Google Scholar WorldCat

McLaughlin, B.P. (forthcoming), 'Could an Android be Conscious?' in A. Pautz and D. Stoljar (eds.), *Themes from Block* (Cambridge, MA: MIT Press).
Google Scholar Google Preview WorldCat COPAC

McLaughlin, B.P., and Hawthorne, J. (1994), 'Dennett's Logical Behaviorism', *Philosophical Topics*, 22: 189–258.
Google Scholar WorldCat

MacPherson (2009), 'Monkey Visual Behavior Falls in "the Uncanny Valley"', *News at Princeton*, October 13.
Google Scholar WorldCat

Mill, J.S. (1865), *An Examination of Sir William Hamilton's Philosophy* (London: Longmans).
Google Scholar Google Preview WorldCat COPAC

Moravec, H. (2003), 'Robots', *Encyclopedia Britannica Online* (July). Retrieved from www.frc.ri.cmu.edu/~hpm/project.archive/robot.papers/2003/robotics.eb.2003.html.
WorldCat

Moravec, H. (2009), 'Rise of the Robots—The Future of Artificial Intelligence', *Scientific American*, March 23.
Google Scholar WorldCat

Moravec, H. (2016), 'Robots', *Encyclopedia Britannica Online*. Retrieved from www.britannica.com/technology/robot-technology.
Google Scholar Google Preview WorldCat COPAC

Mori, M. (1970/2012), 'The Uncanny Valley' (K.F. MacDorman and N. Kageki Trans.), *IEEE Robotics & Automaton Magazine*, 19(2): 98–100.
Google Scholar WorldCat

Nagel, T. (1974), 'What Is It Like to Be a Bat?', *The Philosophical Review*, LXXXIII: 435–50.
Google Scholar WorldCat

Nichols, S., and Knobe, J. (2007), 'Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions', *Noûs*, 41(4): 663–85.
Google Scholar WorldCat

Phelan, M., and Buckwalter, W. (2012), 'Analytical Functionalism and Mental State Attribution', *Philosophical Topics*, 40(2): 129–53.
Google Scholar WorldCat

Saygin, K.P. (2012), 'The Thing That Should Not Be: Predictive Coding and the Uncanny Valley in Perceiving Human and Humanoid Robot Actions', *Social Cognitive Affective Neuroscience*, 7: 413–22.
Google Scholar WorldCat

Seyama, J., and Nagayma, R.S. (2007), 'The Uncanny Valley: Effects of Realism on the Impression of Artificial Human Faces', *Presence: Teleoperators and Virtual Environments*, 16(4): 337–51.
Google Scholar WorldCat

Shieber, S. (2014), 'There Can Be No Turing-Test-Passing Memorizing Machines', *Philosophers' Imprint*, 14(16): 1–13.
Google Scholar WorldCat

Smith, A., and Anderson, A. (2014), 'Digital Life in 2025: AI, Robotics, and the Future of Jobs', *Pew Research Center*, August 6. Retrieved from www.pewresearch.org.
Google Scholar WorldCat

Sparrow, R. (2007), 'Killer Robots', *Journal of Applied Philosophy*, 24(1): 62–77.
Google Scholar WorldCat

Steckenfinger, S.A., and Ghazanfor (2009), 'Monkey Visual Behavior Falls into the Uncanny Valley', *Proceedings of the National Academy of Sciences of the United States of American*, 106(3), July 19.
Google Scholar WorldCat

Sytsma, J., and Machery, E. (2010), 'Two Conceptions of Subjective Experience', *Philosophical Studies*, 151: 299–327.
Google Scholar WorldCat

Tinwell, A., Grimshaw, M., Abdel Nebi, D., and Williams, K. (2011), 'Facial Expressions of Emotion and Perception of the Uncanny Valley in Virtual Characters', *Computers in Human Behavior*, 27(2): 741–9.
Google Scholar WorldCat

Turing, A.M. (1950), 'Computing Machinery and Intelligence', *Mind*, LIX: 433–66.
Google Scholar WorldCat

Wallach, W., and Allen, C. (2009), *Moral Machines: Teaching Robots Right From Wrong* (Oxford: Oxford University Press).
Google Scholar Google Preview WorldCat COPAC

Yeoman, I., and Mars, M. (2012), 'Robots, Men and Sex Tourists', *Futures*, 44: 365–71.
Google Scholar WorldCat

p. 312 Ziff, P. (1959), 'The Feelings of Robots', *Analysis*, XIX: 241–9. ↳
Google Scholar WorldCat

Notes

1 All of the quotations in this paragraph are taken from ASPCR's websites: www.aspcr.com/index.html, www.aspcr.com/newcss_faq.html, www.aspcr.com/newcss_robots.html, and www.aspcr.com/newcss_cruelty.html.

2 For evidence that people see sentience as bearing on moral status, see Jack and Robbins (2012).

3 It would be justifiable to hold such a robot morally accountable in whatever sense normal, adult human beings can be justifiably held morally accountable. We won't discuss here such thorny issues as whether causal determinism is incompatible with moral agency. (See Nichols and Knobe (2007) for a study on peoples' view about that.)

4 There is talk of robot agency in the literature; see Allen, Varner, and Zinger (2000); Allen, Smut, and Wallach (2005); Sparrow (2007); Wallach and Allen (2009). A robot would be an artificial moral agent. But a robot with the requisite mental abilities would be a moral agent *period*. In the sense of "artificial moral agent" relevant to this chapter, the use of the term "artificial" is like the use of that term in "artificial lighting," and unlike its use in "artificial flower." An artificial flower is not a flower. Artificial lighting is genuine lighting. An artificial moral agent in the sense relevant here is just a moral agent that, in addition, happens to be an artifact; similarly for an artificial moral patient.

5 This is intended as a strong supervenience thesis that holds with conceptual necessity (see McLaughlin 1995).

6 Retrieved from www.frc.ri.cmu.edu/~hpm/project.archive/robot.papers/2003/robotics.eb.2003.html. In Moravec (2016), he predicts human-level intelligent robots by 2050.

7 As has often been pointed out, the term "robot" comes from Karel Ĉapek's 1920 Czech language science fiction play *R.U.R.* In the original Czech, "robata" means "forced labor," and is derived from "rab," which means "slave" (Free Online Dictionary).

8 www.frc.ri.cmu.edu/~hpm/project.archive/robot.papers/2003/robotics.eb.2003.html. We are very pleased to report that this claim does not appear in Moravec (2016).

9 *Star Trek: The Next Generation* is a US science fiction television series which ran from 1987 to 1994, created by Gene Roddenberry. It included as a character an android named Lieutenant Commander Data, noted for, among other things, his (Data is a male android) brilliance.

10 *Jeopardy!* is a US television game show created by Merv Griffin. IBM's computer Watson competed on *Jeopardy!* on February 14–16, 2011, and won the grand championship.

11 An American film directed and produced by Robert Zemeckis and distributed by Warner Brothers.

12 An American film directed by Andrew Adamson and Vicky Jenson and produced by PDI/Dream Works.

13 www.npr.org/templates/story/story.php?storyId=124371580.

14 George Henry Lewes (1875) coined the term "anthropomorphize."

15 To fix our ideas, we can think of normal human-level intelligence as involving the capacity (the ability to acquire the ability) to speak and understand a natural language. Although some robots are now being marketed as speaking several languages (e.g., English, Japanese, and German), no extant robot can actually speak a natural language. Neither can Siri on your iPad. We use "speak a natural language" in a literal sense.

16 Of course, carbon is an ingredient of steel. Let's take it that carbon figures in the robots in question only as an ingredient of their steel skeletal structures. What really matters is whether the would-be seat of mental abilities is silicon-based. We ourselves see no reason whatsoever to doubt that a cyborg, with a body made of silicon and steel, but with a normal

human brain, would have both human-level intelligence and sentience. (See Heubner (2010) for supporting evidence that "the folk" think that too.)

17 But see Ziff (1959) and Long (1994) for arguments that robots couldn't feel because they aren't living beings.

18 For the record, one of us, McLaughlin, favors the view that silicon-based sentient robots are not only nomologically impossible, but also metaphysically impossible. But that will be neither assumed nor argued for here.

19 See McLaughlin (forthcoming) for a defense of the view that human-level intelligence and self-consciousness don't require sentience, so that even if there could not be a sentient silicon-based robot, the issue would remain whether there could be a silicon-based robot with human-level intelligence and self-consciousness. We'll say more about sentience shortly.

20 Sytsma and Machery claim that the folk don't conceive of subjective experiences as experiences that are like something for the subject (2010: 229). They maintain that, instead, "for the folk, subjective experience is closely linked to valence" (2010: 299). They tell us that mental states have valence if and only if they have hedonic value for a subject; that is, if and only if they are, for instance, pleasurable (and so have positive hedonic value) or disagreeable (and so have negative hedonic value) (2010: 299). We don't see how an experience could have hedonic value for a subject without the experience being like something for the subject. But an experience can be like something for a subject without having a hedonic value, unless of course the null-hedonic value is included as a value, in which case subjective experiences all trivially have hedonic values. We can't properly discuss here the case that Sytsma and Machery make for their position. We note, though, that we think that they misinterpret the result they get that people are inclined to attribute seeing red to a robot they describe, Jimmy, but are disinclined to attribute pain to Jimmy. We think that the sense in which respondents think Jimmy can see red doesn't indicate that they think Jimmy has visual experiences, and so subjective experiences in our sense. In one sense of "sees red," seeing red is having a visual experience of red. But in another sense of seeing red, to see red is to relationally see the redness of something. (David Rosenthal drew a similar distinction when commenting on a paper of Machery's at a conference held at the City University of New York.) Seeing red in the latter sense doesn't require having a visual experience of red, or indeed any subjective experience (in our sense) at all. Butterflies can see the yellowness of something. But, as we noted, we seriously doubt that they have subjective experiences of any sort. The reason is that we doubt that it is like anything to be a butterfly. Sytsma and Machery acknowledge that someone might object to their position in this way, by claiming that "sees red" has more than one sense, and that in the sense in which folks think Jimmy sees red, seeing red doesn't require having a subjective experience. But they respond that if "sees red" were ambiguous in this way, we should expect people participating in the study to divide roughly evenly on the question of whether Jimmy sees red, some taking the question one way, others taking the question the other way. But we find no such divide. Most folks think that Jimmy sees red. We think the reason that is so is that the scenario Sytsma and Machery describe involving Jimmy invites (what we've called) the relational reading of "sees red," the reading on which seeing red doesn't require having an experience that is like something. That said, what is relevant to our discussion is Sytsma and Machery's result that people tend to be disinclined to attribute pain to Jimmy. As will be apparent in due course, we've found similar results.

21 We take the apt term "peripheral behavior" from Dennett (1987).

22 Topic-neutral terms include terms such as "cause," "effect," "part," and "whole," vocabulary available to physicalists and dualists alike.

23 We'll take no stand here, however, on whether analytical functionalism is true for mental states such as beliefs, desires, and intentions.

24 We should note that not everyone is convinced that analytical functionalism is false for phenomenal consciousness (see Phelan and Buckwalter, 2012). Here we'll just assume that it is.

25 We assume here that laws of nature are metaphysically contingent.

26 We are uncertain whether Chalmers continues to hold it, given that he now seems to embrace panprotopsychism (Chalmers, 2013). Given his embrace of that view, he should regard it as a wide open question whether a robot made of silicon and steel could be phenomenally conscious. Given panprotopsychism, settling that question will have to await the discovery of proto-mental properties, and the determination of how they can combine to produce phenomenal consciousness. A test case for whether a would-be theory of the proto-mental is correct is whether it counts humans and other animals as phenomenally conscious. Whether it counts certain silicon-based robots as phenomenally conscious is not a good test case; whether silicon-based robots could be phenomenally conscious is spoils for the theory of phenomenal consciousness that is victorious on other grounds.

27 See Dennett's (1994) response to McLaughlin and Hawthorne (1994).

28 Dennett's (1971) original intentional systems theory was a kind of instrumentalistic panpsychism. He later stressed the importance of the presence of "real patterns" (1991). On his view, the real patterns that matter to what mental states a being has are patterns of dispositions to peripheral behavior. Internal factors such as functional organization can be relevant, but only insofar as they are relevant to what patterns of disposition to peripheral behavior the being has. (For further discussion, see McLaughlin and Hawthorne, 1994.)

29 It should be mentioned that if laws linking dispositions to behave with states of phenomenal consciousness are supposed to be fundamental laws, rather than laws that hold in virtue of other laws and conditions, then nomological behaviorism would be a kind of ontological emergentism, with states of phenomenal consciousness ontologically emerging from patterns of dispositions to behave. (Similar remarks apply to nomological functionalism.) For discussion of ontological emergence, see McLaughlin (1992); for further discussion of the status of would-be laws linking dispositions to behave with states of phenomenal consciousness, see McLaughlin (2000).

30 For a discussion of how that could mislead people where the issue of robot phenomenal consciousness is concerned, see McLaughlin (forthcoming).

31 In addition to the fact that insects have very few neurons, there is also a ton of behavioral evidence that they don't feel bodily sensations (see Eisemann, Jorgensen, Merritt, Rice, Cribb, Webb, and Zaluki, 1984).

32 We think that nomological functionalism is false, too. But, as we indicated, we will not argue that here.

33 Compare Dennett's position in Dennett (1978).

34 Those who gave an incorrect answer for one or more CRT items were grouped into "Not Pass," while those who gave correct answers for each CRT item were grouped into "Pass."

35 Mantises have minds, but not phenomenally conscious minds. There is, we think, nothing that it is like to be a mantis. For further discussion of insects, see Eisemann et al. (1984) and McLaughlin (forthcoming).

36 Mill's solution is discussed in more detail in McLaughlin (forthcoming).

37 Studies by Phelan and Buckwalter (2012) support, we believe, the view that people rely on folk psychological platitudes connecting beliefs, desire, and the like with states of phenomenal consciousness in attributing the latter states, even when they take themselves to be imagining disembodied individuals. We take that as a reason (though defeasible reason) to think that physical embodiment isn't conceptually required for having phenomenal consciousness.

38 Since one would want to conduct the study using subjects without any background in the philosophy of mind or cognitive psychology, we think it will be difficult to construct a proper questionnaire. The reason is that it would have to convey the technical idea of folk-psychological functional isomorphism, and the fact that the relevant functional organization is describable in physical and topic-neutral terms. From our experiences in teaching undergraduate philosophy of mind courses, that is no easy matter. The idea of two beings with the same dispositions to behave is much easier to get across.

39 McLaughlin would like to thank his former teachers Paul Ziff and Douglas Long for discussions in the late 1970s (his graduate student days) about whether robots could feel, and Christopher Hill (his first philosophy teacher) for the suggestion that an X-phi study would be apt here.