Sandesh Paudel

Dustin Hannum

December 11, 2015

WRT 105

Extending Rights and Moral Considerations To Conscious Artificial Intelligence

Developing rights and moral considerations for machines might seem like a bizarre endeavor. After all, modern-day machines are merely tools designed for human use. Why should machines with artificial intelligence be treated any differently? Granting rights to artificial intelligence might even seem as absurd a proposition as granting rights to blow dryers, toasters and other household appliances. However, the issue at hand might not be as ludicrous as it might seem on surface. There are many consequential reasons to take the idea of granting rights and moral considerations to robots seriously. With the rate of modern advancements in technology and the growth of biological understanding of the human brain and its cognitive processes, it is plausible that we might one day build sophisticated artificial intelligence that has genuine consciousness. This paper aims to explore the moral implications of that possibility, and further discuss how our society ought to respond to conscious robotic entities. If an advanced artificial intelligence can allow a machine to be as cognitively and emotionally sophisticated as humans, then it stands to reason that this machine would deserve the same rights and considerations that we would grant a human.

The question of whether artificial intelligence, in future, would deserve rights, largely remains an untrodden territory. While there are much discussions about programing a sense of "morality" into robots so that they do not harm humans, the potential of artificial intelligence to

act as *recipients* of moral considerations is largely ignored. We tend to be preoccupied with ensuring that robots behave in ways that benefit humans. But how should we humans treat *them*? To give a robot rights is to consider it a "moral patient", and to acknowledge that it has "inherent worth". As John Basl puts it in *Machines as Moral Patients We Shouldn't Care About (Yet): The Interests and Welfare of Current MachinesMark*, "To be morally considerable is to have a welfare composed of interests that are to be taken into account in moral deliberations for the sake of the individual whose welfare it is" (2). Coeckelbergh, a philosopher of media and technology, argues for taking robotic welfare into account by granting inalienable rights to robots when he writes:

> "if (in the future) it turns out that robots share features with humans such as rationality or consciousness, then if we hold these features as a basis for human rights, why restrict those rights to humans? If they might one day become sentient, then why neglect their interests in avoidance of suffering? Why continue to treat artificially intelligent robots as things we can use or abuse if we have good reasons to include them in our community of moral consideration and rights?" (Coeckelbergh, 211).

Coeckelbergh raises the crucial point that if robots achieve the same level of rationality and conscious experiences that humans posses, then we ought to treat them with the same level of dignity and respect as we treat other humans. Coeckelbergh's argument for treating robots as "moral patients" rests on two implicit assumptions (209). The first assumption is that consciousness and rationality are not limited solely to biological beings. These attributes are not "special", in the sense that, they are not exclusive to biological beings made up of flesh and

blood like humans. Under this assumption, a mechanical system with the right components and the right configurations could give rise to a consciousness just as well as the human brain can. It is even conceivable that a robotic consciousness could surpass the conscious abilities of the human mind. The second assumption is that technology will continue to improve in the future. This implies that humanity will build ever more complex machines with faster and better capabilities compared to the previous generation of machines. Eventually, technology will necessarily have to reach a point where it can give rise to consciousness. While the very concept of a conscious machine *seems* like a scenario out of science fiction, the assumptions that guarantee the rise of artificial intelligence are relatively plausible. For a scenario that seems like a far-fetched improbability, there are only *two* necessary assumptions that make it inevitable. Given the rate of technological growth today, and the increase in understanding of the human brain, one can extrapolate that if such progress continues in the future, conscious robots are plausible as a clear possibility. Therefore, the topic of how our society ought to treat robots morally is a very relevant area of discourse.

Coeckelbergh's argument implies that the level of rights and moral considerations that a robot deserves, is intrinsically linked with the features such as rationality and consciousness that the robot "shares with humans" (211). This introduces a simple moral reasoning that applies to robots; the notion that the more a robot mimics human traits, the greater the extent of rights and moral consideration it deserves. We have used this reasoning in various degrees when dealing with animal-rights. In *Extending Legal Rights to Social Robots*, Kate Darling points out that it is only very recently that our legal system has "begun to extend what can be viewed as "second-order" rights to non-human entities"(2), such as animal rights. Moreover, Darling states

that “this effect increases when animals exhibit behavior that we more readily associate with human cognition or emotions” (2). Darling affirms that humans have a demonstrable tendency to treat animals with human-like qualities with more empathy. Darling argues that the same logic would be true for advanced robots (2). As technology progresses, robots would be capable of human-like intelligence, emotional capabilities and other conscious states of mind. Darling asserts that we should grant these robots with appropriate degrees of rights based on the extent to which they resemble humans. However, while humans have a *tendency* to grant rights to beings that are human-like in nature, there is no moral justification on why we *ought to* treat human-like entities with any more respect than we treat beings that may not resemble humans. Why should rights be unique to human nature? Peter Singer, a distinguished moral philosopher of animal rights, has dismissed such an argument by describing it as an example of “speciesism” (Singer, 35). It may simply a bias that humans value things are human-like in nature. Both Darling and Coeckelbergh are guilty of treating humans as the quintessence of beings that deserve moral rights. According to them, any beings that deviate from humanoid features should be granted lesser degrees of rights. This demonstrates an inherent bias in their views.

The human/non-human variance may not be relevant to questions regarding rights and morality. A different approach to granting moral rights is the social relationalist view. The social relationalist view takes a utilitarian approach when granting rights to robots because it argues that the extent of moral rights deserved by a robot is dependent on its contribution to its society. Coeckelbergh presents the social-relational dynamics of morality in *Robot rights? Towards a social-relational justification of moral consideration*. As an analogy to the treatment of robots, Coeckelbergh writes, “one may protect a tree in one’s garden because one loves the view, not

because he thinks that the tree is sentient" (214). In this example, the tree is valued very highly because it is enriching human lives by improving the view. Similarly, one may value and thus grant rights to very intelligent robots because he/she appreciates their contribution to society. The social-relationalist view has no basis on the extent to which a robot resembles humans, or even to the extent to which a robot is "sentient" (214). The degree of rights and moral considerations, according to social relationalists, primarily depends on the fruits of the relationship between humans and the robot (212). Yet the social relationist argument is also biased, because it gives value to an entity based on its contribution to *human* society. Social relationists are guilty of viewing questions about rights and morality solely through the human lens. There is no justification for why a being's ability to contribute to human welfare should be the basis for the rights that it deserves. The social relationist view further breaks down when applied to human themselves. For instance, consider the difference between a billionaire philanthropist and a homeless man. The philanthropist presumably has contributed more to the society than compared to the homeless man. Would the philanthropist then be entitled to a greater degree of rights than compared to the homeless man? It is problematic to grant rights based on one's accomplishments. The modern view of rights is that rights should be equal to all people regardless of social status or personal accomplishments. Considering the social relationalist argument from this perspective, the social relationalist view seems like a selfish proposition because society would value an entity solely based on how much the society has benefitted from it.

A truly just and universal set of moral principles may not be biased towards humans. There must be other criteria that must be examined in order to decide the degree of rights and

moral considerations that an entity deserves. According to McNally in *The rights of robots. Technology, culture and law in the 21st century*, robots

" have restricted mobility, must be artificially programmed for "thought,"lack senses as well as the emotions associated with them, and most importantly cannot experience suffering or fear, they, it is argued, lack the essential attributes to be considered alive. However, the robot of tomorrow will undoubtedly have many of these characteristics and may perhaps become an intimate companion to its human counterpart" (119).

Given the sophistication of future robots, McNally argues that robots will one day have rights. McNally believes that one day, humans, "may see robots in their own right, not only as our mechanical slaves, not only as our products, as ours to buy and sell, but also entities in their own right" (122). McNally ultimately points to characteristics like the ability to experience emotions, suffering and fear, as the basis for a being's ability to act as moral patients. These attributes are not biased towards humans and have the potential to be universal criteria for rights and moral principles. McNally sees the boom in robotics and artificial intelligence to be an opportunity to broaden our and our understanding of rights by extending them to robots and other entities that have the ability to experience a variety of conscious states (119). Granting rights to robots will be the next step in understanding what rights are, in relation to conscious abilities.

In *Artificial consciousness and artificial ethics: Between realism and social relationism*, Torrance expands on McNally's interpretation of rights as a consequence of consciousness. Similar to McNally, Torrance asserts that rights are intrinsically linked with consciousness and sentience (Torrance, 1). He further states that our moral responsibilities must be changed to promote the "well-being" of a conscious entity (4). Torrance suggests that the more consciously

developed an entity gets, the greater the degree of rights it deserves. For instance, a human can experience a much greater scope of suffering, pleasure and other conscious states compared to an ant. The range of experiences for a human is several orders of magnitude greater than that of an ant. In terms of morality, humans are therefore more "important" than ants. Thus, humans deserve a greater level of moral consideration. Torrance argues that non-biological conscious entities like artificial intelligence should be treated in a similar fashion. This view that robots deserve rights based on the extent to which they are conscious is called the "realist" position (5). In contrast to Darling's position and Coeckelbergh's social relationalist view, the realist position does not depend on the distinction between human-like and non-human-like beings, and the distinction between entities that have contributed to human society to various degrees, as the basis of rights and moral considerations. It leads to a universal source of rights that is not biased towards human beings. By defining rights as a consequence of one's conscious ability, the realist position expands the modern idea of rights by extending their objective to promote the wellbeing of conscious beings in general, as opposed to the wellbeing of just humans.

The realist position, however, can give rise to several consequences. Firstly, it is not possible to objectively define states of consciousness. There is no current way of proving that anything is conscious. Instead, we merely look for cues and assume that humans are conscious based on the way they behave. We assume, for instance, that cats do not have the same level of conscious development as humans because they do not demonstrate important cognitive abilities like speech, for instance. However, we can assume that a cat feels pain and pleasure and thus deserves some rights to promote its well-being and to avoid its suffering. Nevertheless, we do not have the capabilities to objectively define levels and states of consciousness. The best we can

do is subjectively interpret behaviors to make assumptions about a being's conscious abilities. It will be particularly difficult, therefore, to know when an artificial intelligence develops consciousness or even sentience. We might have to resort to interpretation of the robot's behavior to determine the level of sentience that it posses. Nevertheless, society should respond to conscious artificial entities similar to how we respond to the ethical treatment of animals. If an artificial intelligence demonstrates a level of consciousness that is equivalent to that of a cat, we ought to treat it as being more "important" than a machine that has conscious capabilities of an ant.

Our technological inability to measure levels of consciousness and to define conscious states, however, does not invalidate the realist position. It merely makes the adoption of the realist position and the determinations about levels of consciousness more difficult. It is also possible that by the time we develop robots capable of consciousness, our understanding of brain will have expanded to completely understand and measure levels of consciousness. In this optimistic scenario, all the barriers that would otherwise prevent us from adopting the realist position would be destroyed. If, on the other hand, machine consciousness is achieved before a good understanding of consciousness, we would be forced to use interpretations and assumptions about a machine's conscious abilities to determine its rights. In fact, many already resort to reasonable assumptions of this kind for animals. Numerous vegans will eat oysters while avoiding eating other organisms. This is because oysters do not have a central nervous system and so it is theorized that they are not capable of feeling pain, or experiencing other conscious states like other animals (Gallary). Making such assumptions about robots seems particularly difficult to do. Nevertheless, the realist position would require us to always look for cues of

conscious abilities. For example, if a robot demonstrates a trait like creativity by engaging to write poetry, we can assume that it has conscious abilities that may rival that of a human. Determinations of conscious abilities of this kind can only be vague, at least until the time we attain a sufficient understanding of what physical processes leads to consciousness.

Granting rights and ethical considerations to robot is a crucial step that our society might need to take as a response to conscious robots. I make the case that despite human limitations in objectively defining consciousness, the realist view holds more merit than the social relationist view. While it may be difficult to objectively measure states of consciousness, the social relationist view and the view that rights are exclusive to human-like entities, are far more biased towards humans to justify them. A proper foundation for rights should be universal, and rely on traits like conscious capabilities as the basis of moral considerations. While much is unknown about consciousness, it is clear that understanding consciousness is the next step in understanding what rights are.

Works Cited:

Coeckelbergh, Mark. "Robot rights? Towards a social-relational justification of moral consideration." *Ethics and Information Technology* 12.3 (2010): 209-221. Web. 18 Nov. 2015.

Basl, John. “Machines as Moral Patients We Shouldn’t Care About (Yet): The Interests and Welfare of Current Machines” Springer Science (2013): 1-5. Web. 09 Dec. 2015.

Darling, Kate. “Extending Legal Rights to Social Robots” *University of Miami* (2012): 1-18. Web. 18 Nov. 2015.

Gallary, Christine. "Why Some Vegans Are OK With Eating Oysters” *Food News. The Kitchn.* Web. 05 Dec. 2015.

McNally, P, Inayatullah, S. “The rights of robots. Technology, culture and law in the 21st century.” *Pergamon.* (1988): 119-136. Web. 18 Nov. 2015.

Millner, Jack. "Should Robots Have Human Rights? Act Now to Regulate Killer Machines before They Multiply and Demand the Right to Vote, Warns Legal Expert." *Mail Online*. Associated Newspapers, 20 July 2015. Web. 17 Nov. 2015.

"Robots Could Demand Legal Rights." *BBC News*. BBC, 21 Dec. 2006. Web. 17 Nov. 2015.

Singer, Peter. “A Utilitarian Defense of Animal Liberation” *Environmental Ethics* (2001), 35. Print.

Torrance, Steve. "Artificial consciousness and artificial ethics: Between realism and social relationism." *Philosophy & Technology* 27.1 (2014): 9-29. Web. 18 Nov. 2015.