Example of robot sentience: the ethics of artificial intelligence essay

War, Intelligence

Since the rise of computers, robots and robotics have captured the imagination of mankind. Science fiction has made many observations regarding robots and potential future artificial intelligence; utilizing artificial intelligence (AI) and robotics to examine fundamental questions of humanity and philosophy has been a method that is used by a variety of different science fiction writers. By creating a world in which artificial intelligence can be compared and contrasted to human intelligence, writers and philosophers can answer questions about the nature of humanity, ethics, and the future of the human race as a whole. In a world where robots are used seven days a week, twenty-four hours a day and a variety of different tasks, some of which harm their human operators, is it wise or ethical to allow robots to have rights?

It is important, first, to discuss what a robot is. Lin et al write, “ Thus a robot must have sensors, processing ability that emulates some aspects of cognition, and actuators. Sensors are needed to obtain information from the environment. Reactive behaviors do not require any deep cognitive ability, but on-board intelligence is necessary if the robot is to perform significant tasks autonomously, and actuation is needed to enable the robot to exert forces upon the environment. Generally, these forces will result in motion of the entire robot or one of its elements” (Lin et al. 943). This definition of “ robot” is particularly useful for discussion because it does not preclude the existence of robots that are biological rather than mechanical, or robots which are a combination of the biological and the mechanical.

One of the issues that is commonly discussed when it comes to artificial intelligence is the issue of sentience. Even today, there are many different

robots that do not have what could be defined as sentience; these robots are designed to do something, and programmed to follow a certain set of protocol. An excellent example of this is a robotic vacuum cleaner; these vacuums are robotic because they act independently of human action, but they cannot be considered sentient. These robots will never suddenly become self-aware, because they are simply not designed or programmed to attain that level of complexity.

However, there do exist robots that are much more complex than robotic vacuums, with the potential to make decisions independent from any outside input from a human. These types of robots raise the question of sentience; when does a robot become sentient? Many people believe that sentience is tied closely to a robot's ability to be self-aware.

Isaac Asimov, a famous science fiction writer, was a proponent of this train of thought; in his novels that dealt with artificial intelligence, sentience was closely tied with self-awareness and a desire for free will. Asimov famously dictated three laws of Machine Ethics in his novel Bicentennial Man: essentially, machines cannot harm humans, a machine must obey human orders as long as it does not interfere with rule number one, and a machine must protect its own life as long as that protection does not interfere with rule one or two (Anderson 1).

Some argue that these rules are good for robots and if humanity ever develops sentient beings, humanity should adopt a similar code of conduct for their robots. However, there is a significant argument against this both morally and ethically. Anderson writes:

Asimov rejected his own Three Laws as a proper basis for Machine Ethics. He

believed that a robot with the characteristics possessed by Andrew, the robot hero of the story, should not be required to be a slave to human beings as the Three Laws dictate. He, further, provided an explanation for why humans feel the need to treat intelligent robots as slaves, an explanation that shows a weakness in human beings Because of this weakness, it seems likely that machines like Andrew could be more ethical than most human beings. (Anderson 1)

Whether or not the robots would be more ethical than human beings is not particularly important; what is important is the question of the ethics of restricting the free will of a sentient being in the same way one would restrict the decision-making process of a being incapable of making independent, logical choices.

Asimov himself rejects the rules as too restrictive and ethically and morally wrong, because there is something fundamentally different about a being that is sentient. When a being cannot think for itself or make independent, logical, and ethical decisions, there is no ethical question regarding sentience; however, when it can, restricting its free will is potentially an ethically poor decision. This raises the question both of the ethics of restricting a sentient being’s free will and of properly monitoring the sentient being for the safety of humanity as a whole. How can humanity balance the rights of the being that it created with the safety of humankind? This is a significant ethical issue that needs to be addressed by those responsible for robotics and artificial intelligence.

However, in Bicentennial Man, the ethical robot in question is one of many robots, most of which are intent upon destroying humanity. One ethical robot

is not enough to make up for the legions of violent, destructive robots that exist in the metaphorical world; freedom of choice means that these robots can choose to be evil, violent, or otherwise destructive. When that happens, humanity's duty is to remove the threat and shut down the issues that the robots are causing.

If a robot has been programmed to be sentient, then humanity must address the underlying ethical question in regards to that robot's free will. Without free will, humanity will have essentially created slaves, which is a concept that has been largely done away with. Some might argue that robots would be more like pets than slaves, but this does not hold out; dogs and cats cannot hold a conversation, for instance, or make complex ethical decisions; robots that have achieved sentience would be able to do both.

This also begs the question of whether or not a robot that is programmed by humanity would ever be able to achieve true sentience-- that is, logical and ethical decision-making skills and self-recognition without being specifically programmed to make those decisions. Any robot that can achieve sentience will also have the ability to learn, and will, more than likely, be much more intelligent than its creators very quickly. Mankind may try to program a kill switch of some kind into this kind of being for fear that the robot may turn on humanity. If they do this, however, it raises the ethical question of whether or not a sentient being should be able to be extinguished so quickly and easily with the press of a button.

Some may argue that sentient robots should never have rights in human society that threaten humanity's rights. This argument has an underlying, arrogant assumption: that these sentient beings would want to be part of

humanity at all. It is equally likely that, given the opportunity, these robots would separate themselves from humanity or declare that humanity is actively hostile to their wants and needs and go to war.

If a sentient robot becomes hostile towards humanity and actively harms humankind, humankind must then make a legal and moral decision regarding whose responsibility that robot is. If the robot is truly sentient, then holding the creator responsible for any harm done seems to be illogical; at the same time, these beings would be made, not born, so anyone harmed by a sentient robot would be harmed indirectly by the creator of that robot. Lin et al. write:

As robots become more autonomous, it may be plausible to assign responsibility to the robot itself consider that synthetic biology is forcing society to reconsider the definition of life, blurring the line between living and non-living agents Also consider that there is ongoing work in integrating computers and robotics with biological brains A conscious human brain (and its body) presumably has human rights, and replacing parts of the brain with something else, while not impairing its function, would seem to preserve at least some of those rights and responsibilities (Lin et al. 946)

Lin et al. suggest that the creators of the robots and the robots themselves should, perhaps, share responsibility for any kind of error or harm done as a result of the robot, but they also bring up an interesting issue: the issue of partially-human robots. At what point does a robot become more human than robot, or vice versa? When does the line between biology and technology blur to a point that it becomes nearly indistinguishable? Perhaps the best discussion of this issue comes from a modern television

series entitled Battlestar Galactica. Battlestar Galactica, a television series that began in 2004, follows humanity through a catastrophic attack of cyborgs (IMDB). These robots, called “ cylons” in the show, were created by humanity to act as servants. However, they became sentient and rebelled against humanity, eventually disappearing into space. The original cylons look very similar to the stereotypical robot: large, silver, and bipedal, they are nearly indestructible.

However, the interesting part of the discussion begins early in the show; the viewer quickly comes to find that the cylons that they recognize as robots are not the only cylons on the show. Some of the cylons appear to be human; indeed, some of the cylons do not even realize that they are not human, having been planted into the human race before the catastrophic attack on humanity to act as sleeper agents if any of humanity escaped (IMDB). This echoes the question raised by Lin et al.: at what point does a robot stop being a synthetic form of life and become human?

When a robot is human in everything but name alone-- the cylons in Battlestar Galactica, for instance, are indistinguishable from humanity except that they do not age and they cannot die (they resurrect in a new, identical body)-- do they cease to be robotic? This is one of the central questions posed by Battlestar Galactica. During the course of the show, humans and cylons fell in love; cylons were betrayed by humans, humans were betrayed by cylons, and both committed acts of violence against the other. Their interactions were marked by suspicion and dislike, and neither side felt any need to trust the other. In short, their interaction was very much like two warring factions of humanity.

Perhaps humanity’s distrust of cylons was understandable. The cylons had just attacked humanity and forced it into exile; as Anderson writes, “ Also in the Twentieth Century, Tibor Machan maintained that to have rights it was necessary to be a moral agent, where a moral agent is one who is expected to behave morally. He then went on to argue that since only human beings possess this characteristic, we are justified in being speciesists: [H]uman beings are indeed members of a discernibly different species – the members of which have a moral life to aspire to and must have principles upheld for them in communities that make their aspiration possible. Now there is plainly no valid intellectual place for rights in the non-human world, the world in which moral responsibility is for all practical purposes absent” (Anderson). Machan would suggest that in the case of the cylons, humanity was right to be suspicious if not downright speciesist; however, morally and ethically, this does not seem to be a logical way to think about the issue. The example of the cylon is a good thought experiment when considering the ethical implications of humanity and robots. The humanoid cyborgs are indistinguishable from humanity, but are still treated with contempt and distrust because many of them have been programmed for aggression against humanity. These cylons often feel badly about their actions in the show, but cannot fight their programming; at one point, a well-liked cylon character has her programming engage and she shoots and nearly kills a human being. Knowing that these cylons can be programmed with hidden agendas is one of the defining moral questions of the show, as it poses the issue of whether or not these creatures can be truly sentient and truly human without complete control over their actions.

Ethically, in most cases, humanity should take its own survival more seriously than the survival of another species or race, particularly one that is bent on destroying humanity. If robots become sentient and cannot be trusted, then they should be disabled and destroyed in any way possible, even if-- or perhaps especially if-- humanity is responsible for their destructive attitudes.

Given this discussion, how likely is it that humanity will ever be able to create a non-human sentient entity? Is all this discussion merely a thought experiment, useful when it comes to questioning ethical standards but fundamentally baseless? Many people do not think so; indeed, many suggest that humanity's ability to create truly sentient life is not as far in the future as many may think. Grossman writes: " We will successfully reverse-engineer the human brain by the mid-2020s. By the end of that decade, computers will be capable of human-level intelligence In that year [2045], he estimates, given the vast increases in computing power and the vast reductions in the cost of same, the quantity of artificial intelligence created will be about a billion times the sum of all the human intelligence that exists today" (Grossman). If Grossman is correct, humanity's ability to create a sentient being is not as far in the future as one may have thought. Science has long had the tendency to experiment openly with things that it may not have control over, with little thought to the long-term consequences; while ethics in science have gotten much better over the years, there are still serious ethical questions to answer before science can create any kind of synthetic being that can think for itself.

This question of the morality of creation is not, of course, a new one. In

Frankenstein, Mary Shelley examines the issue of creating a sentient being when Dr. Frankenstein creates the monster; his monster quickly goes on a rampage, becoming highly intelligent and highly dissatisfied with its life very quickly (Shelley). Frankenstein's monster issues a famous statement to its creator, blaming the doctor for his actions, saying that the Doctor created him in bad faith, with no intention of taking responsibility for the creation of the monster. It is heavily implied that, due to incredible loneliness, the monster intends to kill itself at the end of the novel.

The monster also, like the cylons, notes that it can feel love, and that it desires companionship. Many science fiction writers suggested that the ability to love is the metaphorical " line in the sand" between humanity and robots-- that robots would never be able to truly love, and that this was the way that humanity could draw a distinction between the artificial and the real. However, a truly sentient being would be able to love and fear; it would be able to be hurt and to be angry. Knowing this, it seems as though there are fewer and fewer distinctions that can be drawn between humanity and other sentient beings. This is troubling for humanity, because it means that if-- or when-- humanity creates sentient beings, there will be some very pressing, difficult questions regarding the nature of humanity and what makes a creature human.

Is loneliness the fate that awaits sentient beings created by humanity? One of the biggest problems that faces humanity is a feeling of intense loneliness and isolation. Ironically, this feeling of isolation and of being alone is one of the things that links humanity together as a whole. Would humanity be so cruel as to impose loneliness on sentient beings that they create, to create a

sense of solidarity with them? If the beings that humanity creates are better than humanity itself, would there be jealousy between the two races? Would humanity ever be able to keep a race of sentient beings arguably more intelligent than itself repressed? One would think that the answer is no-- a race of sentient beings that is more intelligent than humanity as a whole would never accept slavery to humankind.

Answering the ethical questions in regards to the rights of robots is important before these beings are created. Too often, humanity jumps into things without fully understanding the outcome. If these beings are created, the issue will no longer be a thought experiment utilized by science fiction writers. Instead, it will be a very real moral, ethical, and safety issue for humanity as a whole.

Humanity is flawed and delicate in many ways, and if it manages to create a race of sentient beings, it is likely that these creatures would be flawed in many of the same ways that humanity is. Pride, ego, fear, jealousy, anger and shame are all emotions that human beings feel in spades; these feelings are often the ones that begin wars and cause problems for humanity. If humanity creates a race of super-powered beings that feels these same feelings, the problems could be incredibly catastrophic for humanity and for robots.

Instinctually, it seems as though creating a race of sentient beings only to enslave them and force them to live as subservient to humanity is not only morally and ethically wrong, but also dangerous. A sentient being will never be happy living a subservient lifestyle that it has not chosen for itself. Compounded with this, it is morally wrong to enslave someone or something

that can think for itself.

The suggestion that humanity should adopt Asimov’s three rules for Machine Ethics addresses this very issue. On the one hand, it seems logical that if humanity is interested in preserving itself, it should do so by restricting the abilities of any sentient being it creates. However, the cost of doing this is very high both morally and ethically. It seems as though if this is done, then humanity would be guilty of doing something morally and ethically suspect in the interest of self-preservation, which is a decidedly ignoble action. There is no good answer to this issue, and yet, it is an issue that needs to be solved quickly and decisively, before humanity creates a being that is both robotic and sentient. Artificial intelligence, arguable, stops being artificial when the being can think organically for itself, and this is a reality that is coming closer and closer with each passing year.

However, humanity has a duty to itself first. When robots become a threat to humankind, that is when the ethical drive for robot rights should stop. If robots become sentient and they show hostility towards human beings, there should be no question about whether or not humans should remove them as a threat, however that can most readily happen.

It is difficult to fathom that robots will ever develop to the point where humanity will have to redefine the idea of “ life,” that day is coming ever closer. At one point, humanity could never have imagined the existence of the Internet; today, the Internet is ubiquitous. The rise of computers was as all-consuming and as imperceptible as the rise of robots has been; today, robotics are used in a variety of different circumstances. If humanity were to

become threatened by robots, then humanity would be in trouble indeed, and should take steps to protect itself from any kind of threat.

Works cited

Anderson, Susan Leigh. " Asimov’s “ Three Laws of Robotics” and Machine Metaethics." University of Connecticut, (n. d.): Print.

Benner, Steven A.. " Q&A: Life, synthetic biology and risk." Journal of Biology, 8. 77 (2010): Print.

Grossman, Lev. " 2045: The Year Man Becomes Immortal." Time 2011: Print.

IMDb. " Battlestar Galactica (TV Series 2004–2009)." 2013. Web. 24 Apr 2013. .

Lin, Patrick et al. " Robot ethics: Mapping the issues for a mechanized world." Artificial Intelligence, 175. (2011): 942–949. Print.

Mcmillan, Graeme. " Four Milestones In the Evolution of Artificial Intelligence." 2011: Print.

Shelley, Mary W. Frankenstein. Charlottesville, Va: University of Virginia Library, 1996. Internet resource.

Waldrop, M. Mitchell. " A Question of Responsibility." AI Magazine 1987: Print.