International Baccalaureate

Philosophy Internal Assessment

Core theme- Artificial Intelligence

The stimulus material: A scene from the movie 'Her'

Title: Could a machine ever be considered a person?

Pavel Garik Pražák

Word count: 1923

Title: Could a machine ever be considered a person?
Stimulus material: Quote from the movie Her (2013), directed and written by Spike Jonze

The movie tells the story of Theodore Twombly, an introverted writer, who purchases an Artificial Intelligence system, named Samantha, to help him write. The quotation displayed below is an advertisement for the system he decides to buy.

''An intuitive entity that listens to you, understands you, and knows you. It's not just an operating system, it's a consciousness. Introducing OS ONE - a life changing experience, creating new possibilities.''

To start with, the stimulus material I have chosen for my Internal Assessment shows an advertisement for the artificial intelligence system OS ONE, from the 2013 movie Her, which is described as 'not just an operating system, it's consciousness'. This statement raises the intriguing philosophical issue of Artificial Intelligence and whether it is possible for a machine to think independently and thus be regarded as a conscious being or even a person. Could a machine be thought of as a person if it simply shows signs of human-like intelligence? Or is there more to being a person? In order to come to a conclusion in regards to the possibility of an artificial intelligence being considered a person, I will firstly scrutinize the term personhood and explore the criteria a being would have to fulfill in order to be regarded as a person. Furthermore, I will discuss and attempt to examine various arguments from several influential philosophers including Alan Turing, John Searle, Mary Ann Warren and Thomas Nagel.

To start with, Mary Ann Warren, an American philosopher, believes that there are six different criteria a being has to attain in order to be considered a person. These include consciousness (including the ability to feel pain), reasoning, self motivated activity, the capacity to communicate, the presence of self awareness and moral agency. While she lists these criteria to express her views on the issue of abortion, they can be used to determine whether a machine could fulfill them and thus be considered a person. According to the

criteria created by Warren, it is obvious that a machine could never be considered a person. While a strong artificial intelligence, meaning a system that could functionally replicate the intellectual capability of a human being, would fulfill many of these criteria such as reasoning and capacity to communicate, it would never be able to replicate the concept of consciousness as such a thing simply can not be recreated. Consciousness is a trait only biological beings possess and creating it artificially is virtually impossible as each being's consciousness is unique. Machines might be able to successfully demonstrate independent thought and intelligence, however, they will never be able to experience the consciousness of a human being. This claim can be supported by Thomas Nagel's famous thought experiment named 'What's it like to be a bat?'. In his paper he states that although we can imagine what it feels like "to be nocturnal, to have webbing on our arms, to be able to fly, to have poor vision and perceive the world through high frequency sound signals, and to spend our time hanging upside down" we will never actually be able to experience what it is like to be a bat as the experience of being a bat can not be replicated. This argument can be used for the issue of artificial intelligence as a computer might be able to replicate almost all aspects of personhood, however, they will never be able to experience what it is like to be a person. Although they will be programmed by persons and have all the properties of one, they will never be able to claim that they are an actual person. Another counter argument against the idea of machines being considered persons comes from religion, specifically Christianity, that claims only a being that has been conceived by two humans and has a soul could be considered a person. This would also dismiss the possibility of a machine being considered a person as it could not possibly be conceived by two human beings and have a soul from god as inanimate objects do not possess such an entity.

On the other hand, computer functionalism, which is a belief that the human mind is an information processing system, and cognition and consciousness constitute a type of computing when combined, claims that we cannot base our belief of who and what we consider a person on the presence of consciousness, soul or other entity that can not be

observed. Instead we should base our conclusions about who we consider to be a person on what is observable. Alan Turing, a world renowned mathematician responsible for breaking the code of the Enigma, believed that a machine should be considered an intelligent being as long as it can not be differentiated from a human being and that it ultimately does not matter what the machine is made out of and whether it has consciousness. We base our perception of people on what we see so why should we take a different approach with computing machines? After all, how can we be sure that other people have consciousness in the first place if it isn't observable? While it is true that consciousness cannot be replicated, all other aspects of personhood could be reproduced by a strong artificial intelligence and since consciousness cannot be seen and we do not consider it when deciding if another human being is a person we should not let the absence of it influence our decision regarding the personhood of a machine. In 1950, Turing proposed a test to determine whether a machine can be considered intelligent. The test would consist of two subjects, one being a human and the other a machine, being questioned by a person who doesn't know who is a machine and who is a human. After the person has interviewed both of the subjects in a written form (so that they can not tell who is a robot by visual and auditory perception) for some time they would have to state which of the two is a machine. If the person's conclusion was incorrect, the machine would pass the Turing test as the interviewer was not able to distinguish between the human and the machine. This would prove that the machine demonstrated human-like behavior and thought and therefore should be considered an intelligent being.

Nevertheless, while Turing's argument suggests that a machine that has successfully passed the test could be considered an intelligent being, there is a strong issue with this claim. John Searle, an American philosopher, has refuted the view that artificial intelligence could ever be considered actually intelligent and capable of independent thought. He did so by introducing a thought experiment known as the 'Chinese room'. In the experiment, a man with no knowledge of the Chinese language is put into a room and given a Chinese rulebook

and dictionaries which tell him how to answer various questions in chinese. Native Chinese speakers then put various questions in Chinese into the room and the man hands them back an appropriate answer, also in chinese. While the people who have put the question into the room now believe the person inside speaks Chinese, he or she actually has no actual knowledge of it whatsoever. What Searle is trying to point out is the fact that although a computer might be able to seem as if it understands us and may appear to be intelligent, the truth is that it just randomly shuffles symbols around in a way it has been programmed to do without any actual intentionality. This would entirely dismiss the view of computer functionalism as the human mind and thought would be impossible to artificially replicate as all computers essentially work by only doing what is written in their code without any actual understanding.

However, Searle's argument lies on the fact that the person in the room does not actually understand Chinese which may be true, nevertheless the Chinese room as a whole system does. While it may have several components that do not understand Chinese independently, they do if they all work collectively. This would mean that while the man in the room does not understand Chinese, the claim that there is no understanding of the questions in Chinese, and that computationalism is false, is denied. Furthermore, Searle didn't necessarily claim that machines are incapable of being intelligent, however, he believed that human intelligence was impossible to recreate as we are made out of biological matter, which computers simply are not and most likely will not ever be. A biological machine that looks and functions like a person might be created and assembled artificially one day, just the way we do now with cars. In such a scenario, we would have succeeded in creating an intelligent machine. Nevertheless, such creation is impossible in today's world meaning any machine is nothing more than an electrical computer encased in a body made of plastic and metal.

Another argument against the idea that a computer could be considered an intelligent being or even a person is the fact that each human being perceives and reacts differently and in

his or her own unique way. However, artificial intelligence systems, no matter how complex, would all most likely use the same line of code meaning their traits would be the same. Humans are genetically very complex and because of it, each person is very different from one another. Some people are naturally more intelligent, sociable and humble while others are unintelligent, introverted and proud. The personality of human beings varies, however, if a strong artificial intelligence was created all systems using it would likely have the same predispositions as they were created by the exact same code. Their personalities would probably develop as they would gain new experiences that would shape the way they perceive and react, nevertheless their initial hardware, programming and experiences would be identical. This would mean that any kind of individuality, that is a trait all people possess, could not be present in a machine unless each code was only used once. However, using the code only once is highly unlikely in a capitalist society as the individual or company that would develop a system capable of human-like behavior and intelligence would likely attempt to generate profit from the creation of such a revolutionary system .

Furthermore, if a machine was to be considered a person it should also have biological and chemical properties a brain of a human has because our behavior and emotions are often just a reaction to a certain chemical being produced in our organs. For example, the feeling of love is in reality just a release of various hormones such as dopamine, oxytocin and serotonin in our brains. With advancements in technology and science, there may be a possibility of creating a biologically artificial brain that would replicate the chemical structure of a person in the future, however, currently no such creation is possible.

In conclusion, while strong artificial intelligence, such as the one presented in the stimulus material, may be able to replicate many of the traits that are typically associated with human beings and attributed to personhood, there are several issues that prevent it from being considered a person. Strong artificial intelligence could certainly not be considered a mindless machine as it would demonstrate many of the characteristics a person possesses

such as intelligence, reasoning, the capacity to communicate etc. Nevertheless, while it will likely have personlike behavior and be almost indistinguishable from an actual human, certain factors such as the absence of biological and chemical properties a human being has, the qualia being the same in multiple systems and the fact that the feeling of being human is impossible to recreate prevent it from being considered a person. To conclude, I believe that the resemblance between a strong artificial intelligence and an actual person will increase over time and will eventually be almost impossible to differentiate, however, completely duplicating a human to the extent the system could be considered a person is impossible. This means that OS ONE, the device that is advertised in the movie 'Her', could never be considered a person because of the reasons mentioned above. However, it could certainly possess a number of human characteristics and could not be treated as an ordinary inanimate object.

References:

Nagel, Thomas. “What Is It Like to Be a Bat?” *The Philosophical Review*, vol. 83, no. 4, [Duke University Press, Philosophical Review], 1974, pp. 435–50, https://doi.org/10.2307/2183914.

Turing, A. M. “Computing Machinery and Intelligence.” *Mind*, vol. 59, no. 236, [Oxford University Press, Mind Association], 1950, pp. 433–60, http://www.jstor.org/stable/2251299.

BYNUM, TERRELL WARD. “ARTIFICIAL INTELLIGENCE, BIOLOGY, AND INTENTIONAL STATES.” *Metaphilosophy*, vol. 16, no. 4, Wiley, 1985, pp. 355–77, http://www.jstor.org/stable/24436830.

“Ethics and Personhood.” *Markkula Center for Applied Ethics*, Santa Clara University, https://www.scu.edu/ethics/focus-areas/bioethics/resources/ethics-and-personhood/.

Dr. Laura D’Olimpio 15 APR 2019, et al. “What Is It like to Be a Bat?” *THE ETHICS CENTRE*, https://ethics.org.au/ethics-explainer-what-is-it-like-to-be-a-bat/.

Van Gulick, Robert. “Consciousness.” *Stanford Encyclopedia of Philosophy*, Stanford University, 14 Jan. 2014, https://plato.stanford.edu/entries/consciousness/.

Cole, David. “The Chinese Room Argument.” *Stanford Encyclopedia of Philosophy*, Stanford University, 20 Feb. 2020, https://plato.stanford.edu/entries/chinese-room/#ReplChinRoomArgu.

Mary Jane Mlynarski “Love, Actually: The Science behind Lust, Attraction, and Companionship.” *Science in the News*, Harvard University, 19 June 2020, https://sitn.hms.harvard.edu/flash/2017/love-actually-science-behind-lust-attraction-companionship/.

On Artificial Intelligence and Consciousness, Journal of Artificial Intelligence and Consciousness, https://www.worldscientific.com/doi/epdf/10.1142/S2705078520500046.