

The Cartesian impossibility for artificially intelligent robots to attain human-level consciousness.

Fatih Incekara

1. Introduction

Although René Descartes is primarily known for his mind-body dualism as a metaphysician, Descartes was more of an anatomist, mathematician, a scientist. Descartes broke with the Scholastic tradition by assigning the lower parts of the Aristotelian tripartite soul to the body itself. He provided detailed neurophysiology including nerves, muscles, and fluids in his *Passion* that had not only implications for moral philosophy, metaphysics, epistemology, and religion, but also for the ‘New Sciences’. Descartes was not only involved in philosophical debates of his time but developed mathematical, anatomical, and technological theories of his own, which influenced many post-Cartesian thinkers such as Baruch Spinoza and Nicolas Malebranche and current scientists within the field of neuroscience and artificial intelligence. Descartes already in the 17th century developed theories about robotics and discussed the possibility of whether robots could attain human-level consciousness (already before pioneers such as Alan Turing). He argued against this possibility because in his age robots were limited in the use of language and were only trained for specific tasks, two characteristics of human-level Cartesian consciousness. With recent technological advances within the field of robotics and artificial intelligence, these assumptions may not hold in our time. Robots now are able to talk and mimic our facial expressions and are able to perform multi-tasks across several domains, which confronts us with the question if these artificially intelligent robots will be able to attain human-level consciousness.

In this essay, I will argue that based on Cartesian arguments, it is impossible for artificially intelligent robots to attain the level of human consciousness. Although Descartes directly argued against this idea already himself, his own arguments do not hold today due to technological advancements. Still, however, using the Cartesian notions of the immaterial, ontological independent mind with *reason, will, and thought* the Cartesian impossibility for artificially intelligent robots to attain the level of human consciousness and experience will be defended.

2.1 Cartesian Consciousness

Aristotle dominated Western philosophy of the soul through thinkers such as Galen and Thomas Aquinas, who consequently influenced thinkers till the seventeenth century such as Descartes. Descartes disliked the notion of the soul and preferred ‘the mind’ which he defined in terms of awareness or consciousness. The notion of the soul was a remnant of the Aristotelians, this scholastic metaphysical tripartite concept of the soul consisted of the vegetative soul, the sensitive soul, and the rational soul. The vegetative soul provided the body with life itself and was also possessed by plants and animals. The sensitive soul was possessed by animals and humans and gave the capacity to sense, feel and accept impressions. The rational soul was human-specific and provided notions such as reasoning, knowledge, and wisdom. Animals did not have any rational souls. Descartes broke with this tradition on purpose as a well-calculated project. The Cartesian mind in contrast does not have any life-giving properties, is an immaterial thinking thing with a reason, a will, and an intellect. It is however not only a thought and a will but also awareness, a consciousness; “*Thought*. I use this term to include everything that is within us in such a way that we are immediately aware of it.”¹ In contrast to this Cartesian mind, the body is exactly what the mind is not; a material nonthinking entity. It is essentially a machine with its own physiology as Descartes describes in his *Passions*, importantly, it is a machine that does not necessarily need a mind to be alive, this is in contrast with the Scholastics. We could speak of a Cartesian trinity, instead of mind-body dualism, however, for Descartes there are three notions; the notion of the mind, the body, *and* the mind-body union. Descartes actually was not necessarily interested in the metaphysical question of how the mind and body causally interact. After correspondence with Princess Elizabeth and her insistence on the explanation about this causality, it is Descartes that proposes a hypothesis that such a causal interaction could interact in the pineal gland of the brain, triggered with the question:

how can the human soul, which is only a thinking substance, determine the movement of the animal spirits in order to perform a voluntary action?”²

Leaving the metaphysical question of causality aside, the notion of the mind-body union has important implications in terms of human experience, awareness, and consciousness. The

¹ AT vii, 160

² AT iii, 661

mind can be influenced by the body, and vice versa. The mind receives impressions (i.e. perception) from the external world through the body and via the five senses. Perception also occurs via information from within the body itself (i.e. the passions proper, both emotions such as love and hate, or feelings such as hunger and thirst):

The first level pertains only to those by which the corporeal organ is immediately affected by external objects. And this can only be the motion of the particles of the organ in question and the change in configuration and position resulting from that motion. The second level contains everything that results immediately in the mind due to the fact that it is united to the corporeal organ so affected, and such are the perceptions of pain, pleasure, thirst, hunger, colors, sound, flavor, smell, heat, cold, and the like, which result from the union and, as it were, an intermingling of mind and body, as I said in *Meditation VI*. The third level comprises all those judgments about external objects which we have been used to making since our earliest childhood on the occasion of the motions of the corporeal organ.³

It should be emphasized that the Cartesian mind is not a Platonic Charioteer, nor an Aristotelian ‘sailor of a ship’ or a Ryleian ‘ghost in a machine.’⁴ The mind is united with the body in a way that, according to Descartes, is metaphysically incomprehensible. But the Cartesian understanding of the mind makes sense of the lived experience; the retina in the eye captures an object with a certain distance, its brown color, its shape, and its sound. Through previously lived experience the mind makes a judgment about this object; it understands that this is a bear coming toward the material body and it makes a judgment with the main goal to preserve the material body. As a result, it becomes aware of the whole situation and moves the body and the limbs away from the object, it thus runs away (or fights) to survive. As a union of the mind with the body, sense perception becomes possible as a kind of thought, such that it can make judgments about the body and the external world possible.⁵

Descartes in a way proposes the proto neuro-anatomical correlates of self-consciousness. He does this in his *Passions*, in which he illustrates that the mind is in need of all the bodily

³ AT vii, 437

⁴ Ryle, Gilbert. 1949. *The Concept of Mind*. London: Penguin Books.

⁵ Rozemond, Marleen. 2006. ‘The Nature of the Mind’ in *The Blackwell Guide to Descartes' Meditations*. Edited by Gaukroger, Stephen. Wiley-Blackwell.

mechanisms, including the brain, nerves, fluids, and all bodily interactions, a process which essentially leads not only to awareness of the external world but also awareness of the body and mind. It becomes aware of emotions such as anger and love. But also to feelings such as thirst and hunger in order to preserve the body by eating and drinking. This is a modern understanding of human experience because its main goal is to provide biological preservation and survival of the body, instead of a more morally laden interpretation of the functions and emotions of the body. Before any religious or metaphysical function of the human mind, it first needs to live and survive in the external world.

The mind and its will however are not necessarily involved in every process to protect the body from danger.⁶ The body can also function without the influences of the mind through reflexes.

When people take a fall and stick out their hands so as to protect their head, it is not reason that instructs them to do this; it is simply that the sight of the impending fall reaches the brain and sends the animal spirits into the nerves in the manner necessary to produce this movement even without any mental volition, just as it would be produced in a machine.⁷

Descartes already distinguishes between the autonomous and somatic nervous systems in the seventeenth century. When mentioning the autonomous nervous system, he notes that “If we think only of enlarging the iris, we can will this as hard as we like, and it makes no difference”.⁸ While he also illustrates the influence of the mind on voluntary movements, “and in general any expression of the face or the eyes can be modified by the soul (...)”⁹ by causing “the gland to drive the spirits to the muscles that produce the desired effect”.¹⁰ As mentioned earlier, a Cartesian proto neuroanatomical correlate of self-consciousness is provided in *the Passions* and in *the Treatise on Man*, not only the pineal gland but the entire Cartesian neurophysiological system including reflexes that is necessary such that perceptions, emotion, and higher cognitive functions can exist. However, this does not imply that the mind *arises* from such biological processes. The mind is in need of the body to understand the external world, but it in principle can exist independently of the body. The

⁶ Hatfield, Gary. 2019. ‘Mind and Psychology in Descartes’ in *The Oxford Handbook of Descartes and Cartesianism*. Edited by Nadler, Steven M, Tad M Schmaltz, and Delphine Kolesnik-Antoine. Oxford: Oxford University Press.

⁷ AT vii, 230

⁸ AT xi, 361

⁹ AT xi, 412

¹⁰ AT xi, 361

same is true for the body. It is in need of the mind to make judgments, to think, to will, but it in principle can be thought of as a body-independent immaterial notion. Both are understood as independent substances. Indeed, Descartes introduced a new ontological independent definition of substance in his *Principles of Philosophy*, in which he defined substance as “a thing which exists in such a way as to depend on no other thing for its existence”¹¹ and a substance can be known from one single principle attribute, which is ‘one principal property that constitutes its nature and essence’.¹² As we will see in the next section, this emphasis is important since advancements in technology raise the question of whether we will be able to build highly complex artificially intelligent robots, hypothetically as sophisticated as humans from the level of a Golgi Apparatus, DNA, a cell, to muscles, organs, nerves, bones and the human brain. Eventually, the question of whether such robots automatically will acquire human-level consciousness needs to be addressed.

2.2 Man, a machine?

Despite the technological advancements in the past decades, our current artificially intelligent robots are still in their early phases of development. These robots include an artificially intelligent algorithm that can learn to speak and perform specific tasks. It is however inevitable that technological advancement will continue and it can be expected that robots will be more complex. When robots will be human-level complex, or to put it more extreme when robots become biological or digital clones of humans (bio-digitally cloned from DNA to the human brain) which may hypothetically happen in the future, will such creatures acquire consciousness, will they acquire human-level consciousness or experience? If this is the case, what is it then, that makes us essentially human? Will we be able to distinguish our child from a digital zombie of our own child? The Turing Test was designed by Alan Turing in 1950, a test in which a robot will pass “the Imitation Game”.¹³ The test is performed in a room in which one human being and one machine have a conversation with an interrogator. Alan Turing’s main objective is to answer the question: Can machines think? If the interrogator can’t tell the difference between the robot and the human, the robot passes the Turing Test; the Robot can think. The Turing test in reality only tells something about the

¹¹ AT viii, 24

¹² AT viii, 25

¹³ Bringsjord, Selmer and Govindarajulu, Naveen Sundar.2019. ‘Artificial Intelligence’, *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2019/entries/artificial-intelligence/>.

human language mimicking ability of the artificially intelligent robot. It does not automatically follow that a robot passing a Turing Test necessarily also achieves a human-like level of consciousness or human experience. Already 300 years before Alan Turing however, Descartes raised the discussion of whether machines will be able to be human-like. Descartes proposes the argument of speech and the argument of task performance:

If there were machines which bore a resemblance to our body and imitated our actions as far as it was morally possible to do so, we should always have two very certain tests by which to recognize that, for all that, they were not real men. The first is that they could never use speech or other signs as we do when placing our thoughts on record for the benefit of others. For we can easily understand a machine's being constituted so that it can utter words, and even emit some responses to action on it of a corporeal kind, which brings about a change in its organs; for instance, if it is touched in a particular part it may ask what we wish to say to it; if in another part it may exclaim that it is being hurt, and so on. But it never happens that it arranges its speech in various ways, in order to reply appropriately to everything that may be said in its presence, as even the lowest type of man can do. And the second difference is, that although machines can perform certain things as well as or perhaps better than any of us can do, they infallibly fall short in others, by which means we may discover that they did not act from knowledge, but only for the disposition of their organs. For while reason is a universal instrument which can serve for all contingencies, these organs have need of some special adaptation for every particular action. From this it follows that it is morally impossible that there should be sufficient diversity in any machine to allow it to act in all the events of life in the same way as our reason causes us to act.¹⁴

The skepticism of Descartes about the possibility that machines will be as real as men essentially arise from the technological limitations of his time. Indeed, the machines in Descartes's time may be limited in the production of speech, in which they only can 'utter words'. In addition, they may only be able to perform specific tasks and 'fall short in others' due to a lack of knowledge. However, recent technological advancements and those in the future very well may be able to build a polyglot artificially intelligent robot that can speak fluently, more fluent in more languages than a human being ever can learn throughout its life. It may be able to master the language with all its cultural nuances and use this during a

¹⁴ AT vi, 66

conversation with a real human. Nor is it unthinkable that future technologies would be able to build a robot that outperforms humans in tasks across many domains, exactly because it is loaded, maybe not with reason, but with knowledge based on billions of pieces of data. It is not unimaginable that an artificially intelligent algorithm of “AlphaGo” for playing the board game “Go” better than humans or the algorithm “Deep Blue” can beat world champion chess player Kasparov in a chess game, the Boston Dynamic robot-soldier and thousands of specifically trained artificially intelligent algorithms would be combined in one single, universal robot. Yes, the ‘Master’ algorithm today does not exist, but its existence in the future is not unthinkable. In other words, these two Cartesian arguments may fall in the future due to the further advancement of robot technologies.

Despite this possibility, Descartes still has a strong argument between the lines of the previous text why indeed it might be impossible for artificially intelligent robots, despite the technological advancements, to acquire human-level consciousness. It is the lack of an immaterial Cartesian mind with reason and intellect that prohibits an artificially intelligent robot from acquiring human-level consciousness. As mentioned in the earlier section, for Descartes, the immaterial mind is not a notion that arises from the complexity of the human body. The mind is an independent substance that exists with the other independent substance, the material body. It is a substance given, and not a substance arising from another substance, which from the Cartesian ontological independence argument is a paradoxical understanding of the relation between the mind and the body, or in the case of an artificially intelligent robot, highly complex, technological and biological system. Therefore, even if science and technology would reach such a level that human DNA, organs, muscles, nerves, the brain, and the whole of the body could be built exactly in the same biological way as the original human; still no human experience, no human self-awareness, no human consciousness would arise in this artificially intelligent biological robot. The artificially intelligent biological robot would pass the Turing Test, but become a mindless, a willless, a thoughtless de La Mettrian machine only mimicking a real human. The robot would be doomed never to be able to have an experience, a self-awareness, or a self-consciousness. Descartes thought that animals are mere automated bodies without a mind, without a capacity to reason and to think. The robot would be doomed, despite its complex architecture, hypothetically more complex than a real animal and as complex as a real human body, to live a life as a Cartesian animal.

3. Conclusion

In this essay, I have argued that based on Cartesian arguments, it is impossible for artificially intelligent robots to acquire human-level consciousness. Descartes already argued against such a possibility based on the argument of speech and the argument of task performance, in which he underestimated the potential of technology. Adhering strictly to these two arguments would fail to support the impossibility argument, especially in light of recent technological advancements. However, despite these recent technological advancements and those potential ones in the future, it will even be impossible for human biologically identical artificially intelligent robots to acquire human-level consciousness, because they do not possess the immaterial, ontological independent Cartesian mind with a reason, will, and thought. Based on the ontological independent argument of a substance, which is given, and does not arise from matter, it is impossible for robots to acquire an immaterial mind. Without a mind, they cannot acquire self-consciousness, nor do they possess the possibility to experience themselves or the external world, this is, independent of potential, hypothetical highly complex biotechnological architecture that they might achieve in the future with further advancements of science within the field such as neuroscience, technology, and artificial intelligence.

4. Bibliography

1. Descartes, René, 1983. *Oeuvres De Descartes*, 11 vols., edited by Charles Adam and Paul Tannery. Paris: Librairie Philosophique J. Vrin.
2. Bringsjord, Selmer and Govindarajulu, Naveen Sundar. 2019. 'Artificial Intelligence', *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2019/entries/artificial-intelligence/>.
3. Rozemond, Marleen. 2006. 'The Nature of the Mind' in *The Blackwell Guide to Descartes' Meditations*. Edited by Gaukroger, Stephen. Wiley-Blackwell.
4. Hatfield, Gary. 2019. 'Mind and Psychology in Descartes' in *The Oxford Handbook of Descartes and Cartesianism*. Edited by Nadler, Steven M, Tad M Schmaltz, and Delphine Kolesnik-Antoine. Oxford: Oxford University Press.
5. Ryle, Gilbert. 1949. *The Concept of Mind*. London: Penguin Books.