MORE HUMAN THAN HUMAN[1]: A Philosophical Exploration of Artificial Intelligence

Author: Sara Masciotra-Milstein
Edited by: Kameron Lee

"I think, Sebastian, therefore I am". The words uttered by an android in Ridley Scott's 1982 cult classic *Blade Runner* allude to Rene Descartes' famous axiom "Cogito ergo sum", thus raising the question as to what it means to be conscious within the context of artificial intelligence. If an artificially intelligent android could replicate human consciousness, then the uniqueness of human identity in relation to rationality would be at stake. An examination of Rene Descartes' philosophy, alongside current research in artificial intelligence and consciousness will support the argument that, should androids evolve in this capacity, their minds would be equal to those of human beings, and thus they would enter personhood.

To commence this discussion on artificial intelligence and human consciousness, it is important to firstly examine Rene Descartes' axiom of existence. In order to reach an absolute truth on the substance of the world, Descartes' philosophical journey stems from a system of hyperbolic doubt, whereby he "rejects as absolutely false anything of which [he] could have the least doubt"[2]. He begins by rejecting any knowledge of the material, external world since the senses are capable of deceiving him. From this position however, Descartes notices that while he could doubt everything outside of himself, the very fact that he doubts indicates that he exists as a thinking thing; hence, his famous axiom: "I think, therefore I am"[3]. From this, Descartes concludes that he is "a substance whose whole essence or nature was only to think, and which, to exist, has no need of space nor of any material thing"[4]. By referring to the mind as its

[1] Quote taken from Ridley Scott's *Blade Runner* (1989).
[2] Rene Descartes, *Discourse on Method*, translated and edited by Laurence J. Lafleur, (Indianapolis and New York: Bobbs-Merrill, 1956), Part Four, 33.
[3] Ibid 33.
[4] Ibid.

own substance, Descartes supports the Christian theology he set out to prove: that there exists a dualism between mind and body, allowing him to speak of a physical world made up of material stuff, and another immaterial realm of the soul and God[5]. Descartes' mind-body dualism claims that "the mind and the body enter into causal relations with each other [where] the mind causes things to happen in the body and the body causes things to happen in the mind"[6]. In this way, Descartes' 'pilot in the ship' analogy follows so that "It would not suffice to place [the rational soul] in the human body, as a pilot in a ship, unless perhaps to move its parts...it must be more intimately joined and united with the body in order to have feelings and appetites like ours, and so constitute a real man"[7]. This inevitably leads to the interactionist problem, which will be examined at a later point. For now, it is suffice to say that Descartes' dualism centers on the premise of the existence of the soul, which for him, is synonymous to mind. This begs the question of if one were to artificially create a mind in a machine (artificial intelligence) would this machine be imbued with a soul?

We must consider that Descartes' human mind —soul— is a self-conscious one, able to recognize its own thoughts and existence, which is a clear issue for Artificial Intelligence[8]. Descartes highlights the uniqueness of human souls, as having been created specially by God, by supporting the "traditional doctrine that the soul is essentially active [which] accounts for Descartes' radical distinction between animal and human consciousness. For Descartes, even when animals are awake ([...] conscious), their images automatically do the work of mediating sensory input and muscular output. But all is dark in the cavities of the brain, where it happens"[9].

For humans, consciousness illuminates this thinking process, so that only human beings are aware of what they are

---

[5] Matt Carter. *Minds and Computers: An Introduction to the Philosophy of Artificial Intelligence*, (Edinburgh: Edinburgh University Press, 2007), 4.
[6] Ibid, 5.
[7] Rene Descartes, *Discourse on Method*, Part Four, 59. One that will be further elaborated upon in the later paragraphs of this essay.
[8] One that will be further elaborated upon in the later paragraphs of this essay.
[9] George Macdonald Ross, "Hobbes and Descartes on Language and Consciousness", (*Synthese* 75, no. 2, 1988), 222.

experiencing and can feel pain or pleasure from it[10]; they are self-conscious. In this way, the human mind is special from any other mind in nature; artificially engineering a mind with the capacity of self-consciousness would threaten humanity's uniqueness from a Cartesian perspective.

Descartes' foresees this issue in some sense. He invokes a mechanical philosophy in order to speak of the human being as "a machine created by the hand of God, and in consequence incomparably better designed and with more admirable movements than any machine that can be invented by man"[11]. Descartes argues that since humans are created by God, they are greater than any man-made machine could ever be. However, given current-day technology, it is possible to see that human creations often surpass or enhance[12] those found in nature (or from God), as seen in laser eye surgery, or bionic prosthetic limbs. Regardless, Descartes' envisions two methods of determining whether a machine, possessing a physiological resemblance to human beings and capable of imitating human actions, were human or not, thus proving the uniqueness of human minds — and souls. The first method whereby one would recognize an android would be that "it could never use words or other signs for the purpose of communicating its thoughts to others as we do"[13], meaning that it could not use language itself, without having been programmed to say certain phrases. The second method explains that "although such machines could do many things as well as, or perhaps even better than men[14] they would infallibly fail in certain others [since they do not] act by understanding, but only by the disposition of their organs. For while reason is a universal instrument which can be used in all sorts of situation, the organs have to be arranged in a particular way for each particular action...it is morally impossible that there should be enough different devices in a machine

[10] George Macdonald Ross, "Hobbes and Descartes on Language and Consciousness", 223.
[11] Rene Descartes, *Discourse on Method* , Part Four, 56.
[12] *Blade Runner* deals with androids who are stronger, faster, and generally more capable than their human counterparts, "more human than human" as it were.
[13] Rene Descartes, *Discourse on Method*, Part Four, 57.
[14] Hence the earlier footnote digression on *Blade Runner*'s androids.

to make it behave in all the occurrences of life as our reason makes us behave"[15]. As in the first method, whereby an android would be incapable of internalizing language for its own use, the second method of recognition demonstrates that androids function on particularities. Without reason's universality (given by God), it would be impossible for an android to possess enough programming for every nuance in human reaction and behaviour. In both methods of recognition, the android would fail because it would merely be "acting in accordance to the disposition of its organs"[16]; it could not be said to possess a rational soul, which works in universals and is thusly reserved for human beings.

Reason as "a universal instrument"[17] is therefore the capacity for free will within human beings. For an android, however, these decisions would have to have been installed in its software, thus denying them free will. For Descartes, it would be impossible for an android to possess a rational soul even if it did reach the same level of intelligence as humans. Human beings would still be unique in this regard.

However, there still remains the interactionist problem in Descartes' philosophy. If the immaterial mind and the body (including the material brain) are separate according to his dualism, then his 'pilot in the ship' analogy, whereby the two interact, is nonsensical. Ironically, Descartes' mechanical philosophy of the human being as a machine created by God lead to atheist principles regarding consciousness[18]. To use a secular term instead, the rational soul of which Descartes speaks, can be understood as "human rational architecture [comprised of] thoughts, beliefs, desires, intentions, emotional states, actions, etc. The entire nexus of rationality relations relating these items to one another, and also to sensory input..."[19]. Unlike the soul, human rational architecture can be explained through

[15] Rene Descartes, Discourse on Method, Part Four, 57-58.
[16] Ibid, 59.
[17] Ibid, 57.
[18] Ann Thomson, "Animals, Humans, Machines and Thinking Matter", (*Early Science and Medicine* 15,no.
1⁄2: 2010), 5.
[19] John Pollock, "Philosophy and Artificial Intelligence", (*Philosophical Perspectives* 4: 1990), 461.

materialist philosophies of mind, such as functionalism and computationalism. According to functionalism, a mental state is defined, not by "anything intrinsic to the state but, rather, its function in mediating relations between inputs, outputs and other mental states. Mental states are held to be functional states"[20].
If artificial intelligence fulfills its function, then it can be said to possess mental states, which leads to computationalism, "the view that to have a mind is to instantiate a particular formal system or collection of systems"[21]. For the computationalist, having a mind means to be engaged in computational processes, meaning that a machine could be said to have a mind. If one were to accept these materialist understandings of the mind as a machine, then it stands to reason that the mind cannot be a system unique to human beings, and that therefore, artificial intelligence is not a threat to human uniqueness since the latter does not exist. The ability to reason complexly and abstractly, is already achievable by computers, and even by some animals[22]. However, it is the higher-order cognitive functions — language production, abstract reasoning, language comprehension[23]— that appear as uniquely human and better serve in the discussion of artificial intelligence as a threat to the uniqueness of human consciousness[24]. For instance, Alan Turing's Imitation Game, whereby a human and a machine answer questions provided to them by an interrogator and must each prove they are human, reflects the machine's gap in language abilities. Passing this test indicates a "sufficient condition" for being a thinking thing; the fact that no computer has thus far been able to pass this test demonstrates the gap still between human consciousness (with its unique capacity for

[20] Matt Carter, *Minds and Computers: An Introduction to the Philosophy of Artificial Intelligence,*
(Edinburgh: Edinburgh University Press, 2007), 45.
[21] Ibid, 95.
[22] Ann Thomson, "Animals, Humans, Machines and Thinking Matter", 6.
[23] Although semantics and linguistics are computational tasks, the development of language and its usage
provide the biggest challenge to Artificial Intelligence researchers.
[24] Matt Carter, *Minds and Computers: An Introduction to the Philosophy of Artificial Intelligence,* 110

language production, abstract reasoning, and language comprehension) and artificial intelligence.

Furthermore, Descartes' systemic doubt of all things external to him consequently results in questions regarding the existence of others' minds. While one can use analogy and inference to rationalize that other humans have mental states 'like' one's own, this 'likeness' is vague[25]. A possible conclusion is that all one can say for certain about these mental states is that they "can be mapped onto our own in such a way that if we suppose the corresponding states to be the same, then the other people are for the most part rational"[26]. It follows then that "the concept of a person must simply be the concept of a thing having states that can be mapped onto our own in such a way that if we suppose the corresponding states to be the same, then the thing is for the most part rational"[27], which means that an android can be considered a person insofar as it appropriately mimics the rational architecture of human minds. This is impossible given Descartes' definition of rationality as independent from the disposition of organs; however, this argument is invalidated given his interactionist problem between mind and body. If the brain is the locus of the mind's rationality and can be understood as functionalist or computation list, then here exists a strong claim in favour of artificial intelligence as equal to human consciousness.

If an android were to reach the level of human consciousness necessary to obtain this status of "personhood", it would be concerning human race, not only philosophically speaking in terms of human specialness, but also in terms of survival; androids as machines could surpass human capabilities (if one chooses explore science fiction). What does it mean to possess the level of consciousness allowing one to be labeled as a 'person', and why do artificially intelligent androids fall short of this? Artificial intelligence researchers answer these questions by epitomizing the interconnectedness between consciousness, personal

[25] John Pollock. "Philosophy and Artificial Intelligence", (*Philosophical Perspectives*4: 1990), 462.
[26] Ibid.
[27] Ibid.

identity, and emotions as integral to human consciousness[28], and by extension, personhood. Both theoretical and practical reasoning require self-consciousness in order to "introspect one's own states, reason about them, and use these results in forming beliefs about the world", as well as a deliberation regarding "what will happen to oneself under various circumstances"[29]. Artificial intelligence would therefore need to possess a rich mental state, furnished with emotions, memories, and self-identity in order to fully possess theoretical and practical reasoning. The android would have to be capable of subjective thinking, which would necessitate the ability to create memories (or the implantation of these by the creator), since emotions, which guide subjective reasoning, are linked to episodic memory[30]. If artificially intelligent androids could possess emotions, humans would be more likely to consider an equivalency between their minds.

This would not be the case, however, if androids merely mimicked emotions without feeling them. Here again lies the problem of the inaccessibility of others minds, a consequence of Descartes' hyperbolic doubt. In everyday interactions, one is content to assume that the people around them feel the emotions they portray[31]. The same analogy would need to apply to artificial intelligence, so that humans, judging solely on the observable behaviour of their mechanical counterparts, would have no way of ascertaining whether the latter's emotions were programmed or truly felt. Regardless, the combination of memories, emotions, and self-consciousness would bestow upon androids a personal identity, equal to that of human beings.

It is with this revelation in mind —that both children of God and their machines—could be qualifiable 'persons'[32], that one must examine the possibility of 'android rights'.
If artificially intelligent androids could be implanted with memories,

---

[28] Matt Carter, *Minds and Computers*, 203.
[29] John Pollock. "Philosophy and Artificial Intelligence", 463.
[30] Matt Carter, *Minds and Computers*, 203.
[31] Matt Carter, *Minds and Computers*, 206.
[32] Given hypothetical technological advances capable of implanting memories and the capacity to feel
emotions for Artificially Intelligent androids.

thus allowing them to feel emotions, then they should be referred to as 'persons'. Descartes' methods for recognizing machine from man would be nullified, since the former would be furnished with the necessary emotions and self-consciousness to act in such a way as to perfectly replicate human behaviour. The issue of an 'android soul' is irrelevant given contemporary secular societies, whereby functionalist and computationalist philosophies are better suited with modern neuroscience and psychology. In these ways, it would be necessary, should artificially intelligent androids ever reach this status of personhood, that their rights be protected. Since they would be capable of feeling emotions, they could not be used as slaves, as seen in *Blade Runner* . The idea of 'rights for robots' is already underway, as modern day courts tackle with the notion of sex robots and whether or not users should be able to rape them, as well as the existence of child sex robots[33]. If these robots were capable of emotions, laws would need to be enacted in order to protect them from what would otherwise be considered criminal activities if committed against a human.

While it is currently technologically impossible for androids to reach this level of personhood, the question as to whether or not the specialness of human consciousness is at stake is only an issue if one accepts Descartes' premise that there exists such a thing as a rational soul unique to human beings. The functionalist and computationalist philosophies refute this in arguing that the human mind is only distinct from machines insofar as it is capable of self-consciousness, emotions, and memories, thus amounting to a self-identity. If humans can create artificial intelligence, something potentially equal to or surpassing their own minds, then this is a true testament to the wonders of the human rationality. Artificial intelligence would become an extension of humanity, able to transcend the only thing limiting human potential: death. Here, one can see the salvation of Descartes' immortal soul he so desperately sought to prove; in a modern society, where the old gods of religion have been replaced by the new gods of

[33] Chris Baynes, "Sex Robots that Let Men Simulate Rape Should Be Outlawed, Says Campaigner". *Independent*: September 21, 2017. http://www.independent.co.uk/

science, artificial intelligence becomes the new soul —surpassing its creator, more human than human.

Works Cited

Baynes Chris. “Sex Robots that Let Men Simulate Rape Should Be Outlawed, Says Campaigner”. *Independent*: September 21, 2017. (Accessed: November 12, 2017).

http://www.independent.co.uk/

Bieri, Peter. "Thinking Machines: Some Reflections on the Turing Test." *Poetics Today* 9, no. 1 (1988): 163-86. doi:10.2307/1772893.

Carter, Matt. *Minds and Computers: An Introduction to the Philosophy of Artificial Intelligence.* Edinburgh: Edinburgh University Press, 2007. Ebook.

Descartes, Rene. *Discourse on Method*, translated and edited by Laurence J. Lafleur. Indianapolis and New York: Bobbs-Merrill, 1956.

Ross, George Macdonald. "Hobbes and Descartes on the Relation between Language and Consciousness." *Synthese* 75, no. 2 (1988): 217-29.

http://www.jstor.org/stable/20116528.

Pollock, John. "Philosophy and Artificial Intelligence." *Philosophical Perspectives* 4 (1990): 461-98. doi:10.2307/2214201.

Thomson, Ann. "Animals, Humans, Machines and Thinking Matter, 1690-1707." *Early Science and Medicine* 15, no. 1/2 (2010): 3-37. http://www.jstor.org/stable/20750207.